



Faculdade de Ciência da Informação – FCI
Programa de Pós-Graduação em Ciência da Informação – PPGCInf

**GESTÃO DE DADOS ABERTOS DE PESQUISA EM UNIVERSIDADES: PROPOSTA
DE DIRETRIZES PARA POLÍTICAS INSTITUCIONAIS COM BASE EM TEORIA
FUNDAMENTADA**

Brasília – DF

Dezembro de 2022

ÉRIKA RAYANNE SILVA DE CARVALHO

**GESTÃO DE DADOS ABERTOS DE PESQUISA EM UNIVERSIDADES: PROPOSTA
DE DIRETRIZES PARA POLÍTICAS INSTITUCIONAIS COM BASE EM TEORIA
FUNDAMENTADA**

Tese de doutorado apresentada ao Programa de Pós-Graduação em Ciência da Informação (PPGCIInf) da Universidade de Brasília (UnB) como requisito parcial para obtenção do título de Doutora em Ciência da Informação, sob a orientação do Prof. Dr. Fernando César Lima Leite e coorientação da Dra. Patrícia Rocha Bello Bertin.

Brasília – DF

Dezembro de 2022

FOLHA DE APROVAÇÃO

Título: “Gestão de dados abertos de pesquisa em universidades: proposta de diretrizes para políticas institucionais com base em teoria fundamentada”

Autor (a): Érika Rayanne Silva de Carvalho

Área de concentração: Gestão, Organização e Comunicação da Informação e do Conhecimento

Linha de pesquisa: **Produção, Socialização e Usos da Informação e do Conhecimento**

Tese submetida à Comissão Examinadora designada pelo Colegiado do Programa de Pós-graduação em Ciência da Informação da Faculdade de Ciência da Informação da Universidade de Brasília como requisito parcial para obtenção do título de **DOCTOR** em Ciência da Informação.

Tese aprovada em: 20 de dezembro de 2022.

Presidente (UnB/PPGCINF): Fernando César Lima Leite

Membro Interno (UnB/PPGCINF): João de Melo Maricato

Membro Externo (UFSCAR): Ariadne Chloe Mary Furnival

Membro Externo (IBICT): Luana Farias Salles

Suplente (FCI/UnB): Michelli Pereira da Costa

Em 17/11/2022.



Documento assinado eletronicamente por **Fernando César Lima Leite, Professor(a) de Magistério Superior da Faculdade de Ciência da Informação**, em 11/01/2023, às 10:42, conforme horário oficial de Brasília, com fundamento na Instrução da Reitoria 0003/2016 da Universidade de Brasília.



Documento assinado eletronicamente por **Ariadne Chloe Mary Furnival, Usuário Externo**, em 12/01/2023, às 10:35, conforme horário oficial de Brasília, com fundamento na Instrução da Reitoria 0003/2016 da Universidade de Brasília.



Documento assinado eletronicamente por **Joao de Melo Maricato, Professor(a) de Magistério Superior da Faculdade de Ciência da Informação**, em 17/01/2023, às 13:21, conforme horário oficial de Brasília, com fundamento na Instrução da Reitoria 0003/2016 da Universidade de Brasília.



Documento assinado eletronicamente por **LUANA FARIAS SALES MARQUES, Usuário Externo**, em 31/01/2023, às 16:43, conforme horário oficial de Brasília, com fundamento na Instrução da Reitoria 0003/2016 da Universidade de Brasília.



A autenticidade deste documento pode ser conferida no site http://sei.unb.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **8984449** e o código CRC **920E98C6**.

Referência: Processo nº 23106.138716/2022-10

SEI nº 8984449

Ficha catalográfica elaborada automaticamente,
com os dados fornecidos pelo(a) autor(a)

CC331g Carvalho, Érika Rayanne Silva de
Gestão de dados abertos de pesquisa em universidades:
proposta de diretrizes para políticas institucionais com
base em teoria fundamentada / Érika Rayanne Silva de
Carvalho; orientador Fernando César Lima Leite; co
orientador Patrícia Rocha Bello Bertin. -- Brasília, 2022.
478 p.

Tese(Doutorado em Ciência da Informação) -- Universidade
de Brasília, 2022.

1. Comunicação científica. 2. Ciência aberta. 3. Dados
científicos. 4. Gestão de dados de pesquisa. 5. Diferenças
disciplinares. I. Leite, Fernando César Lima, orient. II.
Bertin, Patrícia Rocha Bello, co-orient. III. Título.

AGRADECIMENTOS

Assim como o salmista (SL. 69:30), expressei minha gratidão a Deus com cânticos e proclamo sua grandeza com ações de graças. Obrigada, Deus, por ter guardado a minha vida durante a pandemia. Sigo olhando para Jesus, autor e consumidor da minha fé (HEBREUS 12:2).

Agradeço à Universidade de Brasília (UnB), pela oportunidade de ensino, pesquisa e extensão concedida a mim (desde a graduação) e a milhares de estudantes. Viva à universidade pública, gratuita e plural!

Aos professores e funcionários do Programa de Pós-Graduação em Ciência da Informação (PPGCinf) da UnB, pela qualidade de ensino e presteza no atendimento que oferecem, respectivamente.

Ao professor Dr^o Fernando Leite, o qual tenho a honra de tê-lo como orientador nos últimos sete anos. Obrigada por contribuir para o meu crescimento como pesquisadora.

À Dr^a Patrícia Bertin, pela orientação dedicada nesta pesquisa. Você é uma profissional admirável e inspiradora.

Ao professor Dr^o Jayme Vilan Filho, por ter me apresentado o mundo da comunicação científica, ainda em minha graduação. Obrigada por me despertar essa paixão.

À professora Dr^a Michelli Costa, pela atenção, paciência e afeto demandado a mim todas às vezes em que precisei. Obrigada por todo o auxílio e incentivo.

Ao professor Dr^o João Maricato, pelos ensinamentos, atenção, simpatia e colaboração sempre presentes.

Às demais professoras que participaram desta banca – Dr^a Ariadne Chloë Furnival e Dr^a Luana Sales – trazendo contribuições imprescindíveis à tese.

A todos os pesquisadores entrevistados nesta pesquisa, pela contribuição científica em tempos de pandemia e de muitas incertezas vivenciadas no país.

Com amor, agradeço a minha família, que sempre acreditou em meu potencial e foi minha base para que hoje pudesse concluir o doutorado. Em especial, aos meus pais: Roberto Carvalho e Vera Lucia, por todo o amor dado a mim, e as minhas

irmãs: Nágila Carvalho e Nayara Carvalho, pelo afeto e cuidado comigo, desde que eu era pequenininha.

A Paulo Queiroz, pela torcida que sempre dizia: “jogue duro na tese”. Obrigada pela presença em todos os momentos, sejam eles felizes, ou não.

A Cristiane Togo, um presente em forma de amizade concedido pela instituição em que exercemos o cargo de bibliotecário, com muito orgulho. Obrigada, Cris, por ter me ouvido com atenção tantas vezes a respeito de tudo, principalmente, nesses últimos anos.

Por fim, agradeço aos meus queridos amigos que caminharam a jornada de doutorado juntamente comigo: Elton Pinto e Henrique Fernandes. Obrigada por dividirem essa rica experiência ao meu lado.

Amo a vida.
Fascina-me o mistério de existir.

Quero viver a magia
de cada instante,
embriagar-me de alegria.

Que importa a nuvem no horizonte,
chuva de amanhã?
Hoje o sol inunda o meu dia.

Helena Kolody

RESUMO

Práticas de gestão de dados de pesquisa tornaram-se imprescindíveis no atual contexto da pesquisa científica, caracterizado, sobretudo, pela produção de dados em larga escala e pela agilidade na comunicação e uso de informações. O estudo realizado teve como objetivo propor diretrizes para políticas institucionais de gestão de dados abertos de pesquisa em universidades. O referencial teórico adotado na pesquisa teve como base estudos no campo da Ciência da Informação que abrangem o contexto da Ciência e suas classificações, do fenômeno da Ciência Aberta, de dados de pesquisa (ciclo de vida e gestão), repositórios institucionais e metadados para a descrição de conjuntos de dados. O pressuposto teórico no modelo conceitual da pesquisa declara que o entendimento sobre as diferenças entre as áreas é imprescindível na condução de práticas institucionais de gestão de dados abertos de pesquisa em universidades. Do ponto de vista metodológico, a pesquisa adotou o pressuposto filosófico ontológico, à luz da estrutura interpretativa construtivista. O estudo é de natureza analítica e lançou mão de estratégia de investigação qualitativa. Com o propósito de identificar práticas de gestão de dados de pesquisa em diferentes áreas, com base na literatura científica, o estudo analisou artigos de periódicos científicos indexados na base de dados Library and Information Science Abstracts (LISA), por meio do método de revisão sistemática da literatura. Para analisar práticas de gestão de dados de pesquisa realizadas por pesquisadores e descrever fatores que influenciam a abertura de dados de pesquisa em diferentes áreas do conhecimento foram realizadas entrevistas semiestruturadas por videoconferência com bolsistas de produtividade em pesquisa - (nível PQ-2) do CNPq das áreas de Sociologia e Química vinculados a universidades federais brasileiras. Os dados obtidos nas entrevistas foram analisados conforme os processos inerentes à aplicação do método da teoria fundamentada, sendo eles: codificação aberta, codificação axial, codificação seletiva e teoria substantiva. Dentre os resultados obtidos no estudo, destaca-se que a ética na pesquisa é o elemento central para a compreensão das práticas de gestão de dados de pesquisa realizadas na área de Sociologia. Já na Química, esse elemento é a comprovação de dados. Ademais, a infraestrutura para a disponibilização de dados abertos é o aspecto mais importante no contexto dos fatores que influenciam a abertura de dados de pesquisa no campo da Sociologia, enquanto em Química, tal aspecto refere-se a vulnerabilidades na exposição de dados com caráter de inovação tecnológica. Com base nos resultados mencionados foi possível elaborar quinze diretrizes para políticas institucionais de gestão de dados abertos de pesquisa em universidades.

Palavras-chave: Comunicação científica. Ciência aberta. Dados científicos. Gestão de dados de pesquisa. Diferenças disciplinares.

ABSTRACT

Research data management practices have become indispensable in the current context of the scientific research, characterized mainly by large-scale data production and agility in the communication and use of information. The study aimed to propose guidelines for institutional policies of open research data management in universities. The theoretical framework adopted in the research was based on studies in the field of the Information Science that cover the context of the Science and its classifications, of the Open Science phenomenon, of research data (life cycle and management), institutional repositories and metadata for the description of datasets. The theoretical assumption in the conceptual model of the research declares that the understanding of the differences between the areas is essential in conducting institutional practices for the open research data management in universities. Methodologically, the study adopted the ontological philosophical assumption, in light of the constructivist interpretive structure. The study is of analytical nature and used qualitative research strategy. In order to identify research data management practices in different areas, based on scientific literature, the study analyzed articles from scientific journals indexed in the Library and Information Science Abstracts (LISA) database using the systematic literature review method. In order to analyse research data management practices carried out by researchers and to describe factors that influence the opening of research data in different areas of the knowledge, semi-structured interviews by videoconference were conducted with research productivity scholarship holders - (PQ-2 level) from CNPq in the areas of Sociology and Chemistry linked to Brazilian federal universities. The data obtained in the interviews were analyzed according to the processes inherent to the application of the grounded theory method, namely: open coding, axial coding, selective coding and substantive theory. Among the results obtained in the study, it is highlighted that research ethics is the central element for understanding of the research data management practices carried out in the field of Sociology. In Chemistry, this element is the proof of data. Furthermore, the infrastructure for the availability of open data is the most important aspect in the context of the factors that influence the opening of research data in the field of Sociology, while in Chemistry, this aspect refers to vulnerabilities in the exposure of data with character of technological innovation. Based on the results mentioned, it was possible to develop fifteen guidelines for institutional policies for the management of open research data in universities.

Palavras-chave: Scientific communication. Open science. Scientific data. Research data management. Disciplinary differences.

RESUMEN

Prácticas de gestión de datos de investigación se han vuelto imprescindibles en el contexto actual de la investigación científica, caracterizado, sobre todo, por la producción de datos a gran escala y por la agilidad en la comunicación y uso de informaciones. El estudio realizado tuvo como objetivo proponer lineamientos para políticas institucionales de gestión de datos abiertos de investigación en universidades. El marco teórico adoptado en la investigación se basó en estudios en el campo de la Ciencia de la Información que abarcan el contexto de la Ciencia y sus clasificaciones, del fenómeno de la Ciencia Abierta de los datos de investigación (ciclo de vida y gestión), repositorios institucionales y metadatos para la descripción de conjuntos de datos. El supuesto teórico en el modelo conceptual de la investigación declara que la comprensión de las diferencias entre las áreas es fundamental en la realización de prácticas institucionales de gestión de datos abiertos de investigación en universidades. Metodológicamente, la investigación adoptó el presupuesto filosófico ontológico, de acuerdo con la estructura interpretativa constructivista. El estudio es de naturaleza analítica y utilizó la estrategia de investigación cualitativa. Con el fin de identificar prácticas de gestión de datos de investigación en diferentes áreas, con base en la literatura científica, el estudio analizó artículos de revistas científicas indexados en la base de datos Library and Information Science Abstracts (LISA), utilizando el método de revisión sistemática de la literatura. Para analizar prácticas de gestión de datos de investigación realizadas por investigadores y describir factores que inciden en la apertura de datos de investigación en diferentes áreas del conocimiento, se realizaron entrevistas semiestructuradas por videoconferencia con becarios de productividad en investigación - (nivel PQ-2) del CNPq de las áreas de Sociología y Química vinculados a universidades federales brasileñas. Los datos obtenidos en las entrevistas fueron analizados según los procesos inherentes a la aplicación del método de la teoría fundamentada, a saber: codificación abierta, codificación axial, codificación selectiva y teoría sustantiva. Entre los resultados obtenidos en el estudio, se destaca que la ética en la investigación es el elemento central para comprender las prácticas de gestión de datos de investigación realizadas en el campo de la Sociología. En Química, este elemento es la prueba de datos. Además, la infraestructura para la disponibilidad de datos abiertos es el aspecto más importante en el contexto de los factores que influyen la apertura de datos de investigación en el campo de la Sociología, mientras que en Química, este aspecto se refiere a las vulnerabilidades en la exposición de datos con carácter de innovación tecnológica. Con base en los resultados mencionados fue posible elaborar quince lineamientos para políticas institucionales de la gestión de datos abiertos de investigación en universidades.

Palabras clave: Comunicación científica. Ciencia abierta. Datos científicos. Gestión de datos de investigación. Diferencias disciplinarias.

LISTA DE ABREVIATURAS E SIGLAS

ABCD	Access to Biological Collection Data
ABNT	Associação Brasileira de Normas Técnicas
AgMES	Agricultural Metadata Element Set
AMA	Agência para a Modernização Administrativa
AMU	Aligarh Muslim University
BDEP	Exploration and Production Data Bank
BOAI	Budapest Open Access Initiative
CAPES	Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
CAS	Chinese Academy of Science
CBPF	Centro Brasileiro de Pesquisas Físicas
CCZero	Creative Commons Zero Waiver
CD	Compact Disc
CE	Comissão Europeia
CEDAP	Centro de Documentação e Acervo Digital da Pesquisa
CEP/CHS	Comitê de Ética em Pesquisa em Ciências Humanas e Sociais
CESSDA	Consortium of European Social Science Data Archives
CF	Climate and Forecast
CIT	Coordenação de Inovação Tecnológica
CMU	Carnegie Mellon University
CNPq	Conselho Nacional de Desenvolvimento Científico e Tecnológico
CNS	Conselho Nacional de Saúde
CSUC	Consorcio de Serveis Universitaris de Catalunya
DataONE	Data Observation Network for Earth
DATASUS	Departamento de informática do Sistema Único de Saúde do Brasil
DC	Dublin Core
DCC	Digital Curation Centre
DDBJ	DNA DataBank of Japan
DDI	Data Documentation Initiative
DFG	German Research Foundation
DIF	Directory Interchange Format
DILETA	Discurso, Identidade e Letramento Acadêmicos
DOI	Digital Object Identifier
DPI	Direitos de propriedade intelectual
DTU	Technical University of Denmark
DVD	Digital Versatile Disce
DwC	Darwin Core

EML	Ecological Metadata Language
ENA	European Nucleotide Archive
ENS	Exact and Natural Sciences
FAIR	Findable, Accessible, Interoperable, Reusable
FAPESP	Fundação de Amparo à Pesquisa do Estado de São Paulo
FGDC	Federal Geographic Data Committee
FGV	Fundação Getúlio Vargas
Foster Plus	Facilitate Open Science Training for European Research
GFBio	German Federation for Biological Data
GLOBE	Global Collaboration Engine
GNPS	Global Natural Products Social Molecular Networking
HD	Hard Disc
IBGE	Instituto Brasileiro de Geografia e Estatística
IBICT	Instituto Brasileiro de Informação em Ciência e Tecnologia
ICPSR	Inter-university Consortium for Political and Social Research
IES	Instituições de Ensino Superior
Inmetro	Instituto Nacional de Metrologia, Qualidade e Tecnologia
INPE	Instituto Nacional de Pesquisas Espaciais
IODP	International Ocean Discovery Program
IPCA	Índice Nacional de Preços ao Consumidor Amplo
Ipea	Instituto de Pesquisa Econômica Aplicada
IVOA	International Virtual Observatory Alliance Technical Specifications
LAI	Lei de acesso à Informação
LIBER	Ligue des Bibliothèques Européennes de Recherche
LIS	Library and Information Science
LISA	Library and Information Science Abstracts
MAST	Museu de Astronomia e Ciências Afins
NIH	National Institutes of Health
NODC	National Oceanographic Data Center
NSF	National Science Foundation
OAIS	Open Archival Information System
ODbL	Open Data Commons Open Database License
ODC-By	Open Data Commons Attribution License
OECD	Organisation For Economic Co-operation And Development
OKF	Open Knowledge Foundation
ONGs	Organizações Não Governamentais
OpenCon LatAm	Conferencia Latinoamericana em Ciencia Abierta
OSI	Open Society Institute

OSU	Oregon State University
PDDL	Public Domain Dedication and Licence
PLoS	Public Library of Science
PQ	Produtividade em Pesquisa (bolsista – CNPq)
PQu	Pesquisador em Química
RCSB	Research Collaboratory for Structural Bioinformatics
RCUK	Research Councils UK
RDM	Research Data Management
RDMLA	Research Data Management Librarian Academy
RIs	Repositórios institucionais
RN	Resolução Normativa
SciELO	Scientific Electronic Library Online
SDMX	Statistical Data and Metadata Exchange
SSH	Social Sciences and Humanities
SWAFS	Science With And For Society
TCLE	Termo de Consentimento Livre e Esclarecido
TU Delft	Delft University of Technology
UC3	University of California Curation Center
UCL RIISG	Research Information and IT Services Group
UCL	University College London
UFPeI	Universidade Federal de Pelotas
UFPr	Universidade Federal do Paraná
UG	Universidade de Gana
UGCS	Sistema de Computação da Universidade de Gana
UGLS	Sistema de Bibliotecas da Universidade de Gana
UnB	Universidade de Brasília
UNE	Universidade de New England
Unesco	Organização das Nações Unidas para a Educação, a Ciência e a Cultura
UNESP	Universidade Estadual Paulista
UNF	Universal Numerical Fingerprint
Unicef	Fundo das Nações Unidas para a Infância
URLs	Uniform Resource Locator
USGS	United States Geological Survey
USP	Universidade de São Paulo

LISTA DE FIGURAS

Figura 1 – Esquema ilustrativo da árvore de Porfírio.....	44
Figura 2 – Componentes da Ciência aberta.....	65
Figura 3 – Taxonomia da Ciência aberta.....	67
Figura 4 – Cinco fases para a abertura de dados.....	92
Figura 5 – Modelo do ciclo de vida de dados de pesquisa formulado pela Universidade de Virginia...	99
Figura 6 – Modelo do ciclo de vida de dados de pesquisa formulado pela DataOne.....	100
Figura 7 – Modelo do ciclo de vida de dados de pesquisa formulado pelo UK Data Archive.....	103
Figura 8 – Modelo do ciclo de vida de dados de pesquisa formulado por USGS.....	104
Figura 9 – Modelo do ciclo de vida de dados de pesquisa formulado por ICPSR.....	107
Figura 10 – Modelo do ciclo de vida de dados de pesquisa formulado pela GFBio.....	109
Figura 11 – Modelo do ciclo de vida de dados de pesquisa formulado pela Jisc.....	112
Figura 12 – Modelo do ciclo de vida de dados de pesquisa formulado pela biblioteca da DTU.....	114
Figura 13 – Modelo do ciclo de vida de dados de pesquisa formulado por Rüegg <i>et al.</i>	116
Figura 14 – Ciclo de vida de um repositório de dados de pesquisa.....	132
Figura 15 – Fluxo de gestão de dados de pesquisa.....	134
Figura 16 – Modelo conceitual da pesquisa.....	152
Figura 17 – Síntese dos procedimentos adotados na revisão sistemática da literatura.....	172
Figura 18 – Procedimentos de análise de dados relacionados às práticas de gestão de dados de pesquisa realizadas por pesquisadores em diferentes áreas.....	178
Figura 19 – Procedimentos de análise de dados relacionados aos fatores que influenciam a abertura de dados de pesquisa em diferentes áreas.....	180
Figura 20 – Categoria Ciência aberta – Sociologia.....	240
Figura 21 – Categoria Ética na pesquisa – Sociologia.....	242
Figura 22 – Categoria Planejamento – Sociologia.....	245
Figura 23 – Categoria Coleta de dados – Sociologia.....	249
Figura 24 – Categoria Descrição de dados – Sociologia.....	252
Figura 25 – Categoria Documentação de dados – Sociologia.....	253
Figura 26 – Categoria Compartilhamento de dados – Sociologia.....	256
Figura 27 – Categoria Preservação de dados – Sociologia.....	259
Figura 28 – Categoria Reuso de dados – Sociologia.....	261
Figura 29 – Categoria Gestão de dados abertos em universidades – Sociologia.....	264
Figura 30 – Ética na pesquisa e sua relação com práticas de gestão de dados de pesquisa	

- Sociologia.....	269
Figura 31 – Categoria Ciência aberta – Química.....	307
Figura 32 – Categoria Planejamento – Química.....	310
Figura 33 – Categoria Coleta de dados – Química.....	312
Figura 34 – Categoria Descrição de dados – Química.....	314
Figura 35 – Categoria Documentação de dados – Química.....	316
Figura 36 – Categoria Compartilhamento de dados – Química.....	318
Figura 37 – Categoria Preservação de dados – Química.....	320
Figura 38 – Categoria Reuso de dados – Química.....	322
Figura 39 – Categoria Gestão de dados abertos de pesquisa em universidades – Química.....	325
Figura 40 – Comprovação de dados e sua relação com práticas de gestão de dados de pesquisa – Química.....	329
Figura 41 – Categoria Fatores estimuladores à abertura de dados de pesquisa – Sociologia.....	359
Figura 42 – Categoria Fatores inibidores à abertura de dados de pesquisa – Sociologia.....	364
Figura 43 – Categoria Favorável à obrigatoriedade – Sociologia.....	370
Figura 44 – Categoria Desfavorável à obrigatoriedade – Sociologia.....	373
Figura 45 – Infraestrutura para a disponibilização de dados abertos e sua relação com fatores que influenciam a abertura de dados – Sociologia.....	376
Figura 46 – Categoria Fatores estimuladores à abertura de dados de pesquisa – Química.....	405
Figura 47 – Categoria Fatores inibidores à abertura de dados de pesquisa – Química.....	411
Figura 48 – Categoria Favorável à obrigatoriedade – Química.....	416
Figura 49 – Categoria Desfavorável à obrigatoriedade – Química.....	419
Figura 50 – Vulnerabilidades na exposição de dados com caráter de inovação tecnológica e sua relação com a fatores que influenciam a abertura de dados: Química.....	421
Figura 51 – Diferenças entre o registro de patentes e a publicação de artigos.....	424

LISTA DE GRÁFICOS

Gráfico 1 – Número de artigos publicados por ano (2015-2020).....	186
Gráfico 2 – Lugar de publicação dos artigos analisados.....	187
Gráfico 3 – Número de artigos publicados em autoria única e em coautoria.....	189
Gráfico 4 - Perdas acumuladas no Orçamento do Conhecimento entre 2014 e 2022.....	381

LISTA DE QUADROS

Quadro 1 - Cinco escolas de pensamento da Ciência Aberta.....	69
Quadro 2 - Formatos de arquivos.....	88
Quadro 3 - Princípios do FAIR.....	118
Quadro 4 - Estrutura conceitual para gestão de dados de pesquisa.....	120
Quadro 5 - Competências dos bibliotecários na gestão de dados de pesquisa.....	124
Quadro 6 - Funções de um sistema de gestão de dados de pesquisa.....	135
Quadro 7 - Campos mínimos de metadados para o registro.....	144
Quadro 8 - Pressupostos filosóficos com implicações para a prática.....	157
Quadro 9 - Artigos selecionados para o estudo.....	169
Quadro 10 - Formulário de análise para a identificação de práticas de gestão de dados de pesquisa realizadas em diferentes áreas.....	171
Quadro 11 - Procedimentos metodológicos da pesquisa.....	181
Quadro 12 - Artigos que abordaram o planejamento da gestão de dados de pesquisa.....	191
Quadro 13 - Artigo que abordou a coleta de dados no contexto da gestão de dados de pesquisa.....	194
Quadro 14 - Artigos que abordaram a prática do processamento técnico de dados de pesquisa.....	196
Quadro 15 - Artigos que abordaram a prática do compartilhamento de dados de pesquisa.....	199
Quadro 16 - Artigos que abordaram a prática da preservação de dados de pesquisa.....	201
Quadro 17 - Artigos que abordaram a descoberta de dados de pesquisa.....	203
Quadro 18 - Artigos que abordaram a prática do reuso de dados de pesquisa.....	204
Quadro 19 - Artigo que abordou a prática do monitoramento e controle de dados de pesquisa.....	205
Quadro 20 - Artigos que abordaram a prática da elaboração de políticas para a gestão de dados de pesquisa.....	206
Quadro 21 - Artigo que abordou a prática de suporte à análise de dados de pesquisa.....	208
Quadro 22 - Artigos que abordaram a prática da realização de treinamento sobre a gestão de dados de pesquisa orientado a profissionais de informação.....	209
Quadro 23 - Artigos que abordaram a prática da realização de treinamento sobre a gestão de dados de pesquisa orientado a pesquisadores.....	210
Quadro 24 - Síntese das práticas de gestão de dados de pesquisa em diferentes áreas.....	212
Quadro 25 - Codificação aberta da entrevista com PS1: práticas de gestão de dados de pesquisa.....	216

Quadro 26 – Codificação aberta da entrevista com PS2: práticas de gestão de dados de pesquisa.....	218
Quadro 27 – Codificação aberta da entrevista com PS3: práticas de gestão de dados de pesquisa.....	220
Quadro 28 – Codificação aberta da entrevista com PS4: práticas de gestão de dados de pesquisa.....	223
Quadro 29 – Codificação aberta da entrevista com PS5: práticas de gestão de dados de pesquisa.....	225
Quadro 30 – Codificação aberta da entrevista com PS6: práticas de gestão de dados de pesquisa.....	228
Quadro 31 – Codificação aberta da entrevista com PS7: práticas de gestão de dados de pesquisa.....	230
Quadro 32 – Codificação aberta da entrevista com PS8: práticas de gestão de dados de pesquisa.....	232
Quadro 33 – Codificação aberta da entrevista com PS9: práticas de gestão de dados de pesquisa.....	234
Quadro 34 – Codificação aberta da entrevista com PS10: práticas de gestão de dados de pesquisa.....	236
Quadro 35 – Síntese da codificação aberta – Sociologia: práticas de gestão de dados de pesquisa.....	237
Quadro 36 – Codificação aberta da entrevista com PQu1: práticas de gestão de dados de pesquisa.....	284
Quadro 37 – Codificação aberta da entrevista com PQu2: práticas de gestão de dados de pesquisa.....	286
Quadro 38 – Codificação aberta da entrevista com PQu3: práticas de gestão de dados de pesquisa.....	288
Quadro 39 – Codificação aberta da entrevista com PQu4: práticas de gestão de dados de pesquisa.....	290
Quadro 40 – Codificação aberta da entrevista com PQu5: práticas de gestão de dados de pesquisa.....	293
Quadro 41 – Codificação aberta da entrevista com PQu6: práticas de gestão de dados de pesquisa.....	295
Quadro 42 – Codificação aberta da entrevista com PQu7: práticas de gestão de dados de pesquisa.....	

de pesquisa.....	297
Quadro 43 – Codificação aberta da entrevista com PQu8: práticas de gestão de dados de pesquisa.....	300
Quadro 44 – Codificação aberta da entrevista com PQu9: práticas de gestão de dados de pesquisa.....	302
Quadro 45 – Codificação aberta da entrevista com PQu10: práticas de gestão de dados de pesquisa.....	304
Quadro 46 – Síntese da codificação aberta - Química: práticas de gestão de dados de pesquisa.....	305
Quadro 47 – Codificação aberta da entrevista com PS1: fatores que influenciam a abertura de dados de pesquisa.....	342
Quadro 48 – Codificação aberta da entrevista com PS2: fatores que influenciam a abertura de dados de pesquisa.....	343
Quadro 49 – Codificação aberta da entrevista com PS3: fatores que influenciam a abertura de dados de pesquisa.....	345
Quadro 50 – Codificação aberta da entrevista com PS4: fatores que influenciam a abertura de dados de pesquisa.....	346
Quadro 51 – Codificação aberta da entrevista com PS5: fatores que influenciam a abertura de dados de pesquisa.....	348
Quadro 52 – Codificação aberta da entrevista com PS6: fatores que influenciam a abertura de dados de pesquisa.....	349
Quadro 53 – Codificação aberta da entrevista com PS7: fatores que influenciam a abertura de dados de pesquisa.....	350
Quadro 54 – Codificação aberta da entrevista com PS8: fatores que influenciam a abertura de dados de pesquisa.....	352
Quadro 55 – Codificação aberta da entrevista com PS9: fatores que influenciam a abertura de dados de pesquisa.....	354
Quadro 56 – Codificação aberta da entrevista com PS10: fatores que influenciam a abertura de dados de pesquisa.....	355
Quadro 57 – Síntese da codificação aberta – Sociologia: fatores que influenciam a abertura de dados de pesquisa.....	356
Quadro 58 – Codificação aberta da entrevista com PQu1: fatores que influenciam a abertura de dados de pesquisa.....	386

Quadro 59 – Codificação aberta da entrevista com PQu2: fatores que influenciam a abertura de dados de pesquisa.....	388
Quadro 60 – Codificação aberta da entrevista com PQu3: fatores que influenciam a abertura de dados de pesquisa.....	389
Quadro 61 – Codificação aberta da entrevista com PQu4: fatores que influenciam a abertura de dados de pesquisa.....	390
Quadro 62 – Codificação aberta da entrevista com PQu5: fatores que influenciam a abertura de dados de pesquisa.....	392
Quadro 63 – Codificação aberta da entrevista com PQu6: fatores que influenciam a abertura de dados de pesquisa.....	394
Quadro 64 – Codificação aberta da entrevista com PQu7: fatores que influenciam a abertura de dados de pesquisa.....	395
Quadro 65 – Codificação aberta da entrevista com PQu8: fatores que influenciam a abertura de dados de pesquisa.....	397
Quadro 66 – Codificação aberta da entrevista com PQu9: fatores que influenciam a abertura de dados de pesquisa.....	399
Quadro 67 – Codificação aberta da entrevista com PQu10: fatores que influenciam a abertura de dados de pesquisa.....	401
Quadro 68 – Síntese da codificação aberta – Química: fatores que influenciam a abertura de dados de pesquisa.....	402

LISTA DE TABELAS

Tabela 1 – Número de artigos publicados nas quinze universidades brasileiras mais produtivas na <i>Web of Science</i> , entre 2013 e 2018.....	38
Tabela 2 – Palavras-chave atribuídas pelos autores.....	190
Tabela 3 – O número e a proporção de documentos de política que mencionaram os principais estágios do ciclo de vida dos dados de pesquisa.....	193

SUMÁRIO

1 INTRODUÇÃO	29
1.1 Objetivos	31
1.2 Justificativa	32
1.3 Contexto da pesquisa	35
<i>1.3.1 As universidades como lócus de produção do conhecimento científico</i>	35
1.4 Apresentação da tese	40
2 FUNDAMENTAÇÃO TEÓRICA	42
2.1 A Ciência e suas classificações	42
<i>2.1.1 Diferenças disciplinares sob o ponto de vista da comunicação científica</i>	48
<i>2.1.2 Diferenças disciplinares em relação aos dados de pesquisa</i>	53
2.1.2.1 Dados de pesquisa no campo das Ciências Sociais	57
2.1.2.2 Dados de pesquisa no campo da Química	59
2.2 Breve contextualização da Ciência aberta	61
<i>2.2.1 Considerações acerca do acesso aberto</i>	71
2.2.1.1 O que é o acesso aberto?	77
2.3 Dados de pesquisa	81
<i>2.3.1 Tipologia de dados</i>	84
<i>2.3.2 Dados abertos</i>	88
<i>2.3.3 Licenças para conjuntos de dados abertos</i>	93
2.3.3.1 Creative Commons	95
2.3.3.2 Open Data Commons	96
2.4 Modelos do ciclo de vida de dados de pesquisa	98
<i>2.4.1 Modelo do ciclo de vida de dados de pesquisa formulado pela Universidade de Virgínia</i>	98
<i>2.4.2 Modelo do ciclo de vida de dados de pesquisa formulado pela DataOne</i>	100
<i>2.4.3 Modelo do ciclo de vida de dados de pesquisa formulado pelo UK Data Archive</i>	102
<i>2.4.4 Modelo do ciclo de vida de dados de pesquisa formulado pelo USGS</i>	104
<i>2.4.5 Modelo do ciclo de vida de dados de pesquisa formulado pelo ICPSR</i>	106
<i>2.4.6 Modelo do ciclo de vida de dados de pesquisa formulado pela GFBio</i>	109
<i>2.4.7 Modelo do ciclo de vida de dados de pesquisa formulado pela Jisc</i>	111
<i>2.4.8 Modelo do ciclo de vida de dados de pesquisa formulado pela biblioteca da DTU</i>	114
<i>2.4.9 Modelo do ciclo de vida de dados de pesquisa formulado por Rüegg et al.</i>	115

2.5 Gestão de dados de pesquisa	117
<i>2.5.1 O papel do bibliotecário de dados na gestão de dados de pesquisa</i>	125
<i>2.5.2 Serviços de gestão de dados de pesquisa em bibliotecas</i>	127
2.6 Repositório de dados de pesquisa	131
<i>2.6.1 Funções e benefícios dos repositórios de dados de pesquisa</i>	133
<i>2.6.2 Tipos de repositórios de dados de dados de pesquisa</i>	137
2.6.2.1 Repositórios institucionais de dados de pesquisa.....	137
2.6.2.2 Repositórios disciplinares de dados de pesquisa.....	138
2.6.2.3 Repositórios multidisciplinares de dados de pesquisa.....	139
2.6.2.4 Repositórios de dados de pesquisa de projetos específicos.....	139
2.6.2.5 Repositórios de dados de pesquisa próprios.....	140
<i>2.6.3 Repositórios de dados de pesquisa no contexto brasileiro</i>	140
2.7 Metadados para dados de pesquisa	143
<i>2.7.1 Categorias de metadados</i>	145
<i>2.7.2 Padrões de metadados</i>	146
3 METODOLOGIA	150
3.1 Modelo conceitual da pesquisa	150
3.2 Caracterização da pesquisa	155
3.3 Procedimentos metodológicos	167
<i>3.3.1 Identificar práticas de gestão de dados de pesquisa realizadas em diferentes áreas, com base na literatura científica</i>	168
<i>3.3.2 Analisar práticas de gestão de dados de pesquisa realizadas por pesquisadores em diferentes áreas</i>	172
<i>3.3.3 Descrever fatores que influenciam a abertura de dados de pesquisa, segundo pesquisadores em diferentes áreas</i>	178
<i>3.3.4 Síntese dos procedimentos metodológicos adotados para cada objetivo específico proposto no trabalho</i>	181
<i>3.3.5 Descrição do pré-teste</i>	181
<i>3.3.6 Considerações éticas</i>	183
4 APRESENTAÇÃO E DISCUSSÃO DOS RESULTADOS	184
4.1 Identificar práticas de gestão de dados de pesquisa em diferentes áreas, com base na literatura científica	185
<i>4.1.1 Número de artigos publicados por ano</i>	185
<i>4.1.2 Lugar de publicação</i>	187
<i>4.1.3 Tipo de autoria</i>	188
<i>4.1.4 Palavras-chave atribuída pelo autor</i>	189
<i>4.1.5 Práticas identificadas</i>	190
4.1.5.1 Planejamento da gestão de dados de pesquisa.....	191
4.1.5.2 Coleta de dados.....	194
4.1.5.3 Processamento técnico de dados.....	195
4.1.5.4 Compartilhamento de dados.....	199
4.1.5.5 Preservação de dados.....	201
4.1.5.6 Descoberta de dados de pesquisa.....	202

4.1.5.7 Reuso de dados de pesquisa	203
4.1.5.8 Monitoramento e controle de dados de pesquisa	205
4.1.5.9 Elaboração de políticas para a gestão de dados de pesquisa	206
4.1.5.10 Suporte à análise de dados	208
4.1.5.11 Treinamento sobre a gestão de dados de pesquisa orientado a profissionais de informação	209
4.1.5.12 Treinamento sobre a gestão de dados de pesquisa orientado a pesquisadores	210
4.2 Analisar práticas de gestão de dados de pesquisa realizadas por pesquisadores em diferentes áreas	213
<i>4.2.1 Codificação aberta - Sociologia</i>	<i>213</i>
4.2.1.1 Codificação aberta da entrevista com PS1	213
4.2.1.2 Codificação aberta da entrevista com PS2	216
4.2.1.3 Codificação aberta da entrevista com PS3	218
4.2.1.4 Codificação aberta da entrevista com PS4	221
4.2.1.5 Codificação aberta da entrevista com PS5	224
4.2.1.6 Codificação aberta da entrevista com PS6	225
4.2.1.7 Codificação aberta da entrevista com PS7	228
4.2.1.8 Codificação aberta da entrevista com PS8	230
4.2.1.9 Codificação aberta da entrevista com PS9	232
4.2.1.10 Codificação aberta da entrevista com PS10	234
4.2.1.11 Síntese da codificação aberta	237
<i>4.2.2 Codificação axial - Sociologia</i>	<i>239</i>
4.2.2.1 Categoria “Ciência aberta”	239
4.2.2.2 Categoria “Ética na pesquisa”	241
4.2.2.3 Categoria “Planejamento”	244
4.2.2.4 Categoria “Coleta de dados”	249
4.2.2.5 Categoria “Descrição de dados”	251
4.2.2.6 Categoria “Documentação de dados”	253
4.2.2.7 Categoria “Compartilhamento de dados”	256
4.2.2.8 Categoria “Preservação de dados”	259
4.2.2.9 Categoria “Reuso de dados”	261
4.2.2.10 Categoria “Gestão de dados abertos de pesquisa”	263
<i>4.2.3 Codificação seletiva e teoria substantiva relacionada à gestão de dados de pesquisa na área de Sociologia</i>	<i>268</i>
4.2.3.1 Ética na pesquisa e sua relação com práticas de gestão de dados de pesquisa	271
<i>4.2.4 Codificação aberta - Química</i>	<i>282</i>
4.2.4.1 Codificação aberta da entrevista com PQu1	282
4.2.4.2 Codificação aberta da entrevista com PQu2	284
4.2.4.3 Codificação aberta da entrevista com PQu3	286
4.2.4.4 Codificação aberta da entrevista com PQu4	288
4.2.4.5 Codificação aberta da entrevista com PQu5	290
4.2.4.6 Codificação aberta da entrevista com PQu6	293
4.2.4.7 Codificação aberta da entrevista com PQu7	296
4.2.4.8 Codificação aberta da entrevista com PQu8	298

4.2.4.9 Codificação aberta da entrevista com PQu9.....	300
4.2.4.10 Codificação aberta da entrevista com PQu10	302
4.2.4.11 Síntese da codificação aberta.....	305
<i>4.2.5 Codificação axial - Química.....</i>	<i>307</i>
4.2.5.1 Categoria “Ciência aberta”	307
4.2.5.2 Categoria “Planejamento”	309
4.2.5.3 Categoria “Coleta de dados”	312
4.2.5.4 Categoria “Descrição de dados”	313
4.2.5.5 Categoria “Documentação de dados”	316
4.2.5.6 Categoria “Compartilhamento de dados”	317
4.2.5.7 Categoria “Preservação de dados”	320
4.2.5.8 Categoria “Reuso de dados”	321
4.2.5.9 Trabalhos publicados	322
4.2.5.10 Categoria “Gestão de dados abertos de pesquisa em universidades”	324
<i>4.2.6 Codificação seletiva e teoria substantiva relacionada à gestão de dados de pesquisa na área de Química.....</i>	<i>329</i>
4.2.6.1 Comprovação de dados de pesquisa e sua relação com práticas de gestão de dados de pesquisa	331
4.3 Descrever fatores que influenciam a abertura de dados de pesquisa, segundo pesquisadores em diferentes áreas.....	341
<i>4.3.1 Codificação aberta - Sociologia.....</i>	<i>341</i>
4.3.1.1 Codificação aberta da entrevista com PS1	341
4.3.1.2 Codificação aberta da entrevista com PS2	342
4.3.1.3 Codificação aberta da entrevista com PS3	344
4.3.1.4 Codificação aberta da entrevista com PS4	345
4.3.1.5 Codificação aberta da entrevista com PS5	347
4.3.1.6 Codificação aberta da entrevista com PS6	348
4.3.1.7 Codificação aberta da entrevista com PS7	349
4.3.1.8 Codificação aberta da entrevista com PS8	350
4.3.1.9 Codificação aberta da entrevista com PS9	352
4.3.1.10 Codificação aberta da entrevista com PS10.....	354
4.3.1.11 Síntese da codificação aberta.....	355
<i>4.3.2 Codificação axial – Sociologia.....</i>	<i>358</i>
4.3.2.1 Categoria “Fatores estimuladores à abertura de dados de pesquisa”	358
4.3.2.2 Categoria “Fatores inibidores à abertura de dados de pesquisa”	363
4.3.2.3 Categoria “Favorável à obrigatoriedade”	370
4.3.2.4 Categoria “Desfavorável à obrigatoriedade”	373
<i>4.3.3 Codificação seletiva e teoria substantiva referente aos fatores que influenciam a abertura de dados de pesquisa em Sociologia.....</i>	<i>376</i>
4.3.3.1 Infraestrutura para a disponibilização de dados abertos e sua relação com a fatores que influenciam a abertura de dados.....	377
<i>4.3.4 Codificação aberta – Química.....</i>	<i>384</i>
4.3.4.1 Codificação aberta da entrevista com PQu1	384
4.3.4.2 Codificação aberta da entrevista com PQu2	386
4.3.4.3 Codificação aberta da entrevista com PQu3	388

4.3.4.4 Codificação aberta da entrevista com PQu4.....	389
4.3.4.5 Codificação aberta da entrevista com PQu5.....	390
4.3.4.6 Codificação aberta da entrevista com PQu6.....	392
4.3.4.7 Codificação aberta da entrevista com PQu7.....	394
4.3.4.8 Codificação aberta da entrevista com PQu8.....	395
4.3.4.9 Codificação aberta da entrevista com PQu9.....	397
4.3.4.10 Codificação aberta da entrevista com PQu10.....	399
4.3.4.11 Síntese da codificação aberta.....	401
<i>4.3.5 Codificação axial – Química.....</i>	<i>404</i>
4.3.5.1 Categoria “Fatores estimuladores à abertura de dados de pesquisa”.....	404
4.3.5.2 Categoria “Fatores inibidores à abertura de dados”.....	410
4.3.5.3 Categoria “Favorável à obrigatoriedade”.....	416
4.3.5.4 Categoria “Desfavorável à obrigatoriedade”.....	418
<i>4.3.6 Codificação seletiva e teoria substantiva referente aos fatores que influenciam a abertura de dados de pesquisa em Química.....</i>	<i>421</i>
4.3.6.1 Vulnerabilidades na exposição de dados com caráter de inovação tecnológica e sua relação com os fatores que influenciam a abertura de dados.....	422

5 DIRETRIZES PARA POLÍTICAS INSTITUCIONAIS DE GESTÃO DE DADOS ABERTOS DE PESQUISA EM UNIVERSIDADES.....	430
5.1 Solicitar ao pesquisador a elaboração de um Plano de gestão de dados em projetos de pesquisa desenvolvidos no âmbito da universidade.....	431
5.2 Prezar pela proteção ética de dados de pesquisa produzidos na instituição.....	432
5.3 Oferecer suporte a pesquisadores para a construção de repositórios/bancos de dados, como alternativa para o compartilhamento e depósito de dados gerados em projetos de pesquisa.....	433
5.4 Utilizar metabuscadores para recuperar e integrar conjuntos de dados e metadados depositados por pesquisadores em outros repositórios/bancos de dados.....	435
5.5 Adotar estratégias de marketing para promover o reuso de dados abertos disponíveis em repositórios e dar visibilidade a pesquisas cujos dados encontram-se fechados, por meio da divulgação de seus metadados.....	435
5.6 Adotar estratégias de preservação de dados a longo prazo.....	436
5.7 Viabilizar o reuso de dados abertos de pesquisa, prezando pelos direitos autorais sobre dados.....	437
5.8 Elaborar uma Política de gestão de dados abertos de pesquisa.....	437
5.9 Implantar serviço de gestão de dados de pesquisa.....	438
5.10 Capacitar equipes de bibliotecários para a atuação em serviços de gestão de dados de pesquisa.....	439
5.11 Fomentar discussões sobre a gestão de dados abertos de pesquisas na comunidade acadêmica.....	440
5.12 Implantar sistema de recompensas para a valorização do compartilhamento de dados abertos de pesquisa.....	441
5.13 Gerir recursos não somente para a divulgação de resultados de pesquisa publicados, mas também de dados de pesquisa produzidos	

na instituição	442
5.14 Gerir dados abertos de pesquisa de modo a considerar diferenças disciplinares	442
5.15 Promover o depósito de dados abertos de pesquisa em repositórios, por meio de vias dialógicas	442
6 CONCLUSÕES	444
6.1 Identificar práticas de gestão de dados de pesquisa realizadas em diferentes áreas, com base na literatura científica	444
6.2 Analisar práticas de gestão de dados de pesquisa realizadas por pesquisadores em diferentes áreas	445
6.2.1 Aspectos convergentes e divergentes em relação às práticas de gestão de dados de pesquisa realizadas por pesquisadores na área de Sociologia e de Química	448
6.3 Descrever fatores que influenciam a abertura de dados de pesquisa, segundo pesquisadores em diferentes áreas	449
6.3.1 Aspectos convergentes e divergentes em relação aos fatores que influenciam a abertura de dados de pesquisa na área de Sociologia e de Química	452
6.4 Diretrizes que podem vir a ser instituídas para promover a abertura de dados de pesquisa em diferentes áreas do conhecimento	453
6.5 Sugestões de estudos futuros	454
REFERÊNCIAS	456
APÊNDICES	476
Apêndice A: Formulário de análise para a identificação de práticas de gestão de dados de pesquisa realizadas em diferentes áreas	476
Apêndice B: Roteiro de entrevista	477

1 INTRODUÇÃO

Novas práticas na condução da pesquisa científica delineiam-se neste século, ocasionando mudanças basilares em relação às formas de produção e comunicação científica. À luz desse contexto, Dunie (2017) relata que a combinação entre o uso intensivo de tecnologias sofisticadas aplicadas à pesquisa, a comunicação imediata entre pesquisadores e os estudos colaborativos desencadearam uma imensurável produção de dados de pesquisa. Esses, por sua vez, desempenham papel estrutural no processo científico, à medida que podem apresentar tempo de vida superior à duração dos projetos de pesquisa que os geraram. É o que argumentam Sayão e Sales (2015) em relação à possibilidade de pesquisadores e outros profissionais reutilizarem tais dados, ainda que projetos e financiamentos para pesquisas tenham sido finalizados.

Semelhantemente, Silva (2019) destaca que avanços tecnológicos têm contribuído para a produção de grandes quantidades de dados no contexto científico, sobre os quais aponta-se a crescente necessidade de organização, preservação e gestão. Por esse ângulo, o autor argumenta que o intercâmbio de dados é imprescindível para o desenvolvimento científico, à medida que o acesso aos dados possibilita aos pesquisadores reutilizá-los, analisá-los e reproduzi-los para diversos fins. De modo abrangente, o avanço científico se consolida pela gestão efetiva de dados gerados durante o ciclo de vida da pesquisa e de seus resultados, de modo a torná-los visíveis, confiáveis, compartilháveis e úteis.

A literatura científica e técnica é expressiva no tocante a modelos do ciclo de vida de dados elaborados por instituições de ensino e pesquisa. De acordo com Sayão e Sales (2015), tais modelos são relevantes por representarem processos realizados sobre dados ao longo de um projeto de pesquisa, com vistas a sua usabilidade otimizada e prolongada. A título de exemplo, pode-se destacar os modelos do ciclo de vida de dados elaborados por *Data Observation Network for Earth* (DataONE), *UK Data Archive*, *Inter-university Consortium for Political and Social Research* (ICPSR), *United States Geological Survey* (USGS) e Universidade de Virginia.

Segundo Surkis e Read (2015), o conceito de ciclo de vida de dados de pesquisa é comumente utilizado para compreender o propósito e o significado da gestão de dados. À vista disso, os autores argumentam que dados de pesquisa devem ser compreendidos sob perspectiva dinâmica, em contraposição a um cenário cujos dados estão dispostos como peças desconexas ao sistema de produção científica. Dessa forma, os dados de pesquisa não estão isolados dos fenômenos investigados e possuem histórias que representam o próprio fazer científico. Em razão disso, é imprescindível que sejam geridos para fins de recuperação e reuso.

Considerando o contexto da *e-Science*, Bertin, Visoli e Drucker (2017) destacam a introdução da gestão de dados de pesquisa como uma prática fundamental no âmbito estratégico de organizações. Logo, é importante que um programa institucional dessa natureza se preocupe em dispor aos pesquisadores diretrizes e instrumentos adequados para a gestão de dados de pesquisa durante todo o ciclo de vida dos mesmos. Para tanto, a primeira medida a ser tomada é situar a realidade dos dados de pesquisa na instituição, ou seja, realizar um diagnóstico sobre esses dados. Sendo assim, é fundamental esclarecer alguns aspectos, como os tipos de dados de pesquisa que a organização lida, o modo pelo qual dados são produzidos/obtidos, validados, armazenados e reutilizados, a preservação de dados a longo prazo, a qualidade e a segurança de conjuntos de dados, entre outros.

Dentre os aspectos relevantes à discussão sobre dados de pesquisa, especialmente no que se refere a sua gestão, é oportuno destacar a noção de diferenças entre áreas, termo relativo à diversidade de práticas entre áreas do conhecimento em relação à produção/obtenção, compartilhamento e uso da informação no processo científico. De acordo com Borgman (2015), tais diferenças refletem-se no contexto dos dados de pesquisa em razão da própria natureza dos domínios de pesquisa, que apresentam especificidades quanto aos objetivos que visam atingir e à adoção de métodos para a coleta e análise de dados.

Diante do cenário delineado, observa-se que a pluralidade das áreas, tal como um fenômeno que se manifesta nos modos pelos quais dados e resultados de pesquisa são produzidos e comunicados, é nitidamente perceptível no ambiente de universidades. Isso, em razão de essas instituições serem espaços tradicionalmente

ocupados por diversas áreas do conhecimento, apesar das particularidades existentes. Aliada a esse contexto, a noção de diferenças entre as áreas é altamente relevante para se conduzir a gestão de dados abertos de pesquisa. Entretanto, tal aspecto constitui um grande desafio vivenciado pelas universidades, refletindo a necessidade da institucionalização da prática de gestão de dados abertos de pesquisa, de modo a considerar a pluralidade das áreas do conhecimento.

À vista disso, observa-se a importância da realização de estudos que se dediquem ao enriquecimento do corpo de conhecimento relacionado à gestão de dados de pesquisa em diferentes áreas do conhecimento no ambiente de universidades. Assim, este estudo pretendeu buscar respostas às perguntas: 1) Quais diretrizes podem vir a ser instituídas para promover a abertura de dados de pesquisa em diferentes áreas do conhecimento? 2) Como tem sido realizada a gestão de dados de pesquisa em diferentes áreas? 3) Quais fatores influenciam a abertura de dados de pesquisa em diferentes áreas?

1.1 Objetivos

Considerando a complexidade inerente ao ambiente de universidades no que concerne a padrões de comportamento de diferentes áreas, o objetivo geral deste estudo é propor diretrizes para políticas institucionais de gestão de dados abertos de pesquisa em universidades. De modo a alcançá-lo, foram definidos três objetivos específicos:

- 1) identificar práticas de gestão de dados de pesquisa realizadas em diferentes áreas, com base na literatura científica;
- 2) analisar práticas de gestão de dados de pesquisa realizadas por pesquisadores em diferentes áreas;
- 3) descrever fatores que influenciam a abertura de dados de pesquisa, segundo pesquisadores em diferentes áreas.

1.2 Justificativa

Apesar do avanço tecnológico e científico logrado ao longo do tempo, alguns desafios conhecidos pela humanidade se mantiveram no presente século, e, ao mesmo tempo, novos surgiram. Sob perspectiva sistêmica e surtindo efeitos que se desdobrarão por gerações, tais desafios demonstram a fragilidade humana em assegurar a preservação do meio ambiente, garantir educação indistintamente, erradicar a pobreza, promover a saúde e a cura de doenças.

A adoção de medidas de prevenção a doenças e a cura de enfermidades tornaram-se ações de extrema necessidade em nível mundial, principalmente em razão da pandemia da Covid-19, deflagrada no ano de 2019, advinda do novo Coronavírus. À vista do exposto, algumas iniciativas centradas em dados abertos de pesquisa têm sido desenvolvidas para contribuir no estudo da COVID-19 e na geração de novas pesquisas.

A título de exemplo, pode-se destacar a criação do repositório *COVID-19 Data Sharing/BR*¹, resultado de esforços da Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) em colaboração com a Universidade de São Paulo (USP), e participação do Instituto Fleury, Hospital Sírio-Libanês e Hospital Israelita Albert Einstein. Esse repositório disponibiliza dados demográficos, dados derivados de exames clínicos e laboratoriais, além de informações sobre o estado de pacientes, incluindo o número de internações e óbitos. Ademais, o conteúdo do repositório é atualizado constantemente e os dados que disponibiliza são pseudonimizados.

Considerando o delineamento de um novo contexto de pesquisa científica caracterizado, sobretudo, pela intensa produção de conhecimentos e pela celeridade no compartilhamento e no reuso de informações devido à necessidade de superar desafios que afligem a humanidade, destaca-se que discussões sobre dados de pesquisa ascendem em momento oportuno.

De acordo com Federer (2016), o pesquisador do século XXI ocupa-se com grandes conjuntos de dados digitais que ele mesmo produz, ou obtém de outras

1 Disponível em: <https://repositoriodatasharingfapesp.uspdigital.usp.br/>

fontes disponíveis. Além disso, cada vez mais, agências financiadoras de pesquisas e editores de periódicos científicos têm requisitado o compartilhamento de dados de pesquisa e a elaboração do Plano de gestão de dados a pesquisadores.

Sendo assim, estudos que se comprometam à reflexão da gestão de dados abertos de pesquisa em diferentes áreas também demonstram sua relevância em nível teórico e prático. Por isso, os parágrafos seguintes desta seção dedicam-se em apresentar as contribuições em nível teórico e prático, as quais este relatório de tese de doutorado pretende oferecer.

Em nível teórico, observa-se que há uma lacuna a ser preenchida no campo da Ciência da Informação, particularmente no contexto da comunicação científica, relacionada ao estudo do fenômeno da gestão de dados abertos de pesquisa, com base em diferentes padrões de comportamento das áreas do conhecimento. Considerando a lacuna apontada, este trabalho reconhece que a universidade é um ambiente bastante rico em termos de pluralidade de áreas e esta noção traz reflexos que incidem fortemente na maneira pela qual a gestão de dados abertos é conduzida em nível institucional.

Do mesmo modo, este trabalho parte do pressuposto de que “se o conhecimento é o motor do avanço científico, os dados são seu combustível”, tal como declara Silva (2019, p. 3). Ainda de acordo com o autor, a gestão de dados de pesquisa tornou-se um dos principais tópicos de discussão no âmbito profissional de diversas áreas, todavia, foi na Ciência da Informação que ocupou espaço de destaque, principalmente em razão da necessidade de se desenvolver novas infraestruturas para assegurar a preservação digital de dados de pesquisa. Para tanto, é imprescindível estabelecer quais dados deverão ser preservados, como os mesmos serão geridos, quais instituições constituirão parâmetro de boas práticas realizadas neste campo, dentre outros aspectos.

Considerando as particularidades inerentes ao ambiente de universidades, os resultados deste estudo pretendem contribuir em nível prático para a gestão de dados abertos de pesquisa em diferentes áreas, oferecendo insumos para o desenho e implantação de políticas institucionais e de sistemas de informação convenientemente direcionados às áreas. Tal aspecto reflete um grande desafio vivenciado pelas

universidades, que por sua natureza, tradição e construção social ao longo do tempo, caracterizam-se sobretudo pela reunião das áreas do conhecimento e, conseqüentemente, lidam com a pluralidade de práticas relacionadas aos modos pelos quais dados de pesquisa são produzidos e comunicados. Esse é o principal motivo que justifica a realização do estudo no contexto de universidades.

Tendo ainda como pano de fundo as contribuições em nível prático as quais este trabalho pretende oferecer, acrescenta-se, por oportuno, comentários tecidos por Silva (2019) sobre aspectos econômicos que influenciam a gestão de dados. Acerca disso, o autor explica que a coleta de dados é uma das atividades mais dispendiosas no processo da pesquisa científica. Por essa razão, revela-se a necessidade de uma gestão eficaz de dados, tendo em vista que muitos dados coletados não são compartilhados em acesso aberto, tampouco organizados em repositórios de dados com vistas ao seu reuso. Nesse mesmo sentido, Sayão e Sales (2015) destacam que dados podem apresentar tempo de vida maior que a duração dos projetos de pesquisa que os produziram e, desse modo, serem reutilizados. Portanto, mesmo após projetos serem finalizados, dados de pesquisa podem ser úteis a outros estudos, sendo combinados a resultados de pesquisas, ou integrados a novos elementos.

À vista do exposto, convém destacar que o estudo enfatizou as perspectivas de pesquisadores nas áreas de Sociologia e de Química, com a finalidade de demarcar claramente as influências das diferenças disciplinares nas práticas de gestão de dados abertos de pesquisa no âmbito de universidades. Tal como será abordado na seção metodológica do presente trabalho, trata-se de um recorte amostral em que as disciplinas mencionadas são representantes de duas grandes áreas: *Social Sciences and Humanities* (SSH) e *Exact and Natural Sciences* (ENS), respectivamente.

Ressalta-se, ainda, que esta pesquisa reconhece a existência de dados não abertos ou que não podem ser abertos por alguma razão. A gestão sobre os mesmos também se demonstra desafiadora, e, ao mesmo tempo, necessária. Porém, conforme o objetivo geral declarado no estudo, foi dada ênfase na gestão de dados abertos de pesquisa quando se propôs diretrizes para políticas institucionais em universidades. Sendo assim, o recorte realizado privilegiou a perspectiva de dados abertos de pesquisa e dos fatores relacionados a sua abertura.

1.3 Contexto da pesquisa

As diretrizes para políticas institucionais de gestão de dados abertos de pesquisa as quais este trabalho se propôs a elaborar dizem respeito ao ambiente de universidades. Considerando que essas instituições apresentam características intrínsecas que se refletem nas funções que exercem enquanto entidades de ensino, aprendizagem, produção de pesquisas científicas e extensão de projetos à sociedade, julgou-se prudente oferecer uma contextualização que aborde aspectos importantes relativos à universidade. Isso porque entende-se que a caracterização do ambiente do estudo é relevante para se obter uma dimensão prévia dos fenômenos, objetos e sujeitos que serão correlacionados no decorrer da pesquisa.

Nessa direção, a próxima seção abordará aspectos importantes acerca do ambiente de universidade, apresentando-o como *lócus* de referência para a produção do conhecimento científico em diversas áreas do conhecimento. Além disso, será abordada a relevância das universidades brasileiras no ciclo produtivo da pesquisa científica.

1.3.1 As universidades como lócus de produção do conhecimento científico

Dentre os possíveis sentidos atribuídos ao termo “universidade”, um deles extraído do “Dicionário Houaiss da língua portuguesa” (HOUAISS; VILLAR; FRANCO, 2009) parece ser o mais apropriado para caracterizar o contexto do qual se trata esta pesquisa. De acordo com o verbete do dicionário referido, pode-se entender universidade como sendo:

Instituição de ensino e pesquisa constituída por um conjunto de faculdades e escolas destinadas a promover a formação profissional e científica de pessoal de nível superior, e a realizar pesquisa teórica e prática nas principais áreas do saber humanístico, tecnológico e artístico e a divulgação de seus resultados à comunidade científica mais ampla (HOUAISS; VILLAR; FRANCO, 2009, p. 1907-1908).

Com base no verbete apresentado, infere-se que a universidade é uma instituição em que se desenvolve a prática do ensino e da pesquisa no âmbito da

educação de nível superior. Logo, pesquisas de natureza teórica e prática são realizadas e disseminadas em diferentes contextos disciplinares, incluindo o campo das Humanidades, da Tecnologia e das Artes.

Acrescente-se, por oportuno, verbete extraído do “Novo Aurélio Século XXI: o dicionário da língua portuguesa” (FERREIRA, 1999), no qual enfatiza-se a função basilar da universidade em promover a preservação e o avanço científico nas diversas áreas do conhecimento, por meio do ensino e da pesquisa. Nesse sentido, a universidade pode ser entendida como *lócus* de referência onde se desenvolve a pesquisa científica e se produz conhecimentos em diferentes ramos, conforme pode-se observar:

[Universidade é a] instituição de ensino superior que compreende um conjunto de faculdades ou escolas para a especialização profissional e científica, e tem por função precípua garantir a conservação e o progresso nos diversos ramos do conhecimento, pelo ensino e pela pesquisa (FERREIRA, 1999, p. 2032).

Segundo Severino (2002), a produção do conhecimento, que é elemento-chave para o destino da humanidade, encontra seu expoente máximo no ambiente de universidade. Sob essa perspectiva, as atividades de pesquisa, extensão e ensino complementam-se entre si de maneira indissociável à universidade. Na pesquisa se aprende e se ensina, e por meio dela, serviços são prestados à sociedade.

Logo, a pesquisa científica no âmbito das universidades assume uma tríplice dimensão. A primeira, a dimensão epistemológica, refere-se à construção do conhecimento e à significação de objetos e fenômenos. A segunda compreende a dimensão pedagógica, a qual está relacionada com os processos de aprendizagem. Por fim, a terceira dimensão diz respeito à dimensão social, que inclui a perspectiva da extensão (SEVERINO, 2002).

A pesquisa produzida no ambiente de universidade é essencial para a extensão de produtos e serviços à sociedade. Por meio da pesquisa, o conhecimento produzido pode ser disseminado e submetido a condições de universalização. Sob esse ponto de vista, o conhecimento é compreendido como um bem simbólico que precisa ser usufruído em benefício de toda a sociedade, por intermédio da universidade (SEVERINO, 2002).

Apesar do importante papel das universidades em relação à produção de pesquisas científicas, conforme apontou Severino (2002), Schwartzman e Castro (1986)

destacam que em determinado momento, o alto custo proveniente da realização de pesquisas associado ao crescente potencial econômico, culminaria por afastar definitivamente a pesquisa do ambiente das universidades. Conseqüentemente, essa atividade estaria concentrada em institutos especializados e seria manejada conforme padrões semelhantes aos adotados no contexto da produção industrial.

Entretanto, a tendência demonstrada parece reverter-se principalmente devido o desenvolvimento de tecnologias. Sobretudo em países desenvolvidos, têm-se observado a crescente importância do ambiente universitário e o surgimento de novas culturas de cooperação científica com outros setores da sociedade, incluindo governos e indústrias. O paradoxo se revela, na medida em que apesar de a pesquisa científica atrair a atenção de governos e decisores políticos, essencialmente, é na universidade que tal atividade ocorre e se desenvolve. Mesmo países desenvolvidos em termos de recursos aplicados à pesquisa tecnológica no âmbito de empresas e institutos governamentais, são as universidades que dispõem da maior quantidade de pesquisadores, e é nessas instituições que se produz a ciência mais básica (SCHWARTZMAN; CASTRO, 1986).

Abordando a pesquisa universitária como um dos principais elementos do sistema científico, Brickman (1986) tece o seguinte comentário:

A pesquisa universitária é um dos principais componentes do sistema científico. Muito do que sabemos sobre o comportamento e as normas dos cientistas provém da análise da ciência universitária, ou, mais especificamente da pesquisa básica praticada nas universidades. Somente à medida que formos aprendendo mais acerca da influência de contextos institucionais e sociais específicos sobre a atividade científica, seremos capazes de melhor distinguir os traços mais gerais da ciência daqueles que ela assume no ambiente acadêmico (BRICKMAN, 1986, p. 37).

No que tange à infraestrutura pública em pesquisa abordada em relatório que demonstra as relações entre universidade e empresas no Brasil, De Negri, Cavalcante e Alves (2013) afirmam que a existência de um vasto e atualizado parque de pesquisa em ciência e tecnologia é fundamental para o crescimento da produção de conhecimentos em um país e um dos requisitos para que haja inovação nessas áreas. Além disso, parcela considerável de recursos investidos em pesquisa e

desenvolvimento em um país, geralmente, procede de universidades e instituições públicas de pesquisa.

A importância das universidades brasileiras no ciclo produtivo da pesquisa científica demonstra-se clarividente no relatório solicitado pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) à empresa *Clarivate Analytics*, intitulado: “*Research in Brazil: Funding excellence Analysis prepared on behalf of CAPES by the Web of Science Group*”. Publicado em 2019, o relatório foi produzido pelo *Web of Science Group* e oferece um panorama atual da pesquisa brasileira no âmbito global, por meio da análise bibliométrica de dados de artigos brasileiros publicados entre os anos de 2013 e 2018. Os dados analisados foram extraídos da base *Web of Science*. A Tabela 1 apresenta o número de artigos em cinco categorias da CAPES (Ciências da Saúde, Ciências Biológicas, Ciências Exatas e da Natureza, Ciências Agrícolas, Engenharia) publicados nas quinze universidades brasileiras que possuem a maior produção indexada na *Web of Science*, entre 2013 e 2018.

Tabela 1 – Número de artigos publicados nas quinze universidades brasileiras mais produtivas na *Web of Science*, entre 2013 e 2018

	Ciências da Saúde	Ciências Biológicas	Ciências Exatas e da Natureza	Ciências Agrícolas	Engenharia	Todas as pesquisas
Universidade de São Paulo	21,912	17,025	14,536	6,476	6,819	58,899
Universidade Estadual Paulista	5,283	6,948	5,336	5,908	2,914	22,868
Universidade Estadual de Campinas	5,719	4,416	6,571	1,989	3,941	19,317
Universidade Federal do Rio de Janeiro	4,672	5,351	5,503	981	3,038	17,484
Universidade Federal do Rio Grande do Sul	5,199	4,009	3,960	2,168	2,599	15,860
Universidade Federal de Minas Gerais	5,233	4,349	3,293	1,809	2,108	14,904
Universidade Federal de São Paulo (UNIFESP)	7,372	3,186	1,212	358	724	11,228
Universidade Federal do Paraná	2,133	3,333	2,486	2,190	1,628	9,995
Universidade Federal de Santa Catarina (UFSC)	2,473	1,974	2,468	1,358	2,284	9,162
Universidade Federal de Pernambuco	1,778	2,302	2,391	662	1,082	7,098
Universidade de Brasília	1,756	2,039	2,023	895	892	7,056
Universidade do Estado do Rio de Janeiro	2,110	1,315	3,046	281	1,030	7,039
Universidade Federal de São Carlos	977	1,727	2,643	670	2,072	6,980
Universidade Federal de Viçosa	602	2,726	940	3,064	441	6,893
Universidade Federal de Santa Maria (UFSM)	1,247	1,809	1,425	2,522	782	6,670

Fonte: Traduzida e adaptada de *Web of Science Group* (2019, p. 24)

Conforme pode ser visto na Tabela 1, a produção das universidades mencionadas corresponde a mais de 50% de toda a produção de pesquisa no Brasil nestas áreas. O que a Universidade de São Paulo (USP) produz sozinha equivale a mais do que o dobro do número total de publicações obtido pela Universidade Estadual Paulista (UNESP), que é a segunda colocada em termos de produção de pesquisa. Acrescente-se, por oportuno, que as quinze universidades referidas são todas universidades públicas (WEB OF SCIENCE GROUP, 2019).

1.4 Apresentação da tese

Para fim de organização dos conteúdos apresentados, esta tese foi segmentada em seis grandes seções que se referem, nesta ordem: à introdução; à fundamentação teórica; à metodologia; aos resultados atingidos; às diretrizes para políticas institucionais de gestão de dados abertos de pesquisa em universidades e às conclusões obtidas no estudo.

Na introdução discutiu-se sobre questões as quais a pesquisa pretendeu responder, os objetivos propostos e as contribuições que o estudo visou oferecer. Assim também, foi caracterizado o contexto onde a pesquisa está situada – as universidades, que, por sua vez, são reconhecidas como importante *locus* para a produção do conhecimento científico.

A fundamentação teórica apresentou estudos obtidos por meio de investigação bibliográfica. Subdividindo-se em vários capítulos, esta seção da tese explorou principalmente as seguintes temáticas:

- perspectiva histórica das classificações do conhecimento e das ciências;
- contextualização da Ciência Aberta;
- dados de pesquisa – aspectos concernentes à tipologia de dados de pesquisa, aos dados abertos, às licenças para conjuntos de dados abertos;
- modelos do ciclo de vida de dados de pesquisa;
- gestão de dados de pesquisa – incluindo estudos sobre o papel do bibliotecário de dados na gestão de dados e a criação de serviços de dados em bibliotecas;
- repositório de dados de pesquisa;
- metadados para dados de pesquisa.

Com relação à seção de Metodologia, nela se descreveu o modelo conceitual adotado, e a pesquisa foi caracterizada em termos de estratégias de investigação e dos procedimentos metodológicos empregados para a obtenção de cada objetivo específico proposto no estudo – a partir da descrição das fontes de dados, das técnicas de coleta e análise de dados.

A seção de resultados reuniu todos os achados da pesquisa, possibilitando a proposição de diretrizes para políticas institucionais de gestão de dados abertos de pesquisa em universidades. Ressalta-se, ainda, que para evitar ambiguidades na interpretação de resultados relacionados ao objetivo de analisar práticas de gestão

de dados de pesquisa realizada por pesquisadores em diferentes áreas e ao de descrever fatores que influenciam a abertura de dados de pesquisa, segundo pesquisadores em diferentes áreas, os resultados obtidos no estudo serão apresentados conforme os processos da teoria fundamentada (codificação aberta, codificação axial, codificação seletiva e teoria substantiva), primeiramente, no contexto da Sociologia, e logo depois, no de Química.

As diretrizes para políticas institucionais de gestão de dados abertos de pesquisa em universidades podem ser consultadas após a apresentação dos resultados do estudo. Por fim, as conclusões atingidas em relação aos objetivos propostos e as sugestões de estudos futuros serão mencionadas na seção de conclusão.

2 FUNDAMENTAÇÃO TEÓRICA

Esta seção apresentará o resultado de investigação bibliográfica na qual foram consultadas fontes de informação presentes na literatura científica e técnica. Abordará os seguintes aspectos: 1) produção e comunicação científica relacionadas às diferenças disciplinares, iniciando a narrativa por estudos acerca das classificações da Ciência; 2) Ciência Aberta; 3) dados de pesquisa, com ênfase na tipologia, nos dados abertos, licenças para conjuntos de dados abertos, modelos do ciclo de vida de dados e na gestão de dados de pesquisa; 4) repositórios de dados de pesquisa; 5) metadados de dados de pesquisa. A investigação realizada foi relevante na identificação de elementos que fundamentaram em nível teórico o estudo, principalmente em relação à elaboração do modelo conceitual adotado na pesquisa.

2.1 A Ciência e suas classificações

Este trabalho estrutura-se sobre o pilar das diferenças entre as áreas e parte do pressuposto que o entendimento sobre tais diferenças é imprescindível para a gestão de dados abertos de pesquisa. Para tanto, isto requer a apreciação do desenvolvimento das próprias áreas do conhecimento, em termos de especialização, ao longo do tempo. Por essa razão, a seção inicia uma breve explanação com cunho histórico a respeito das classificações da Ciência, para logo então, abordar aspectos relevantes sobre diferenças disciplinares sob o ponto de vista da comunicação científica.

De acordo com Pombo (1998), a classificação dos saberes sempre despertou o interesse da humanidade, principalmente no que se refere ao conhecimento científico. Além disso, as classificações das ciências foram historicamente produzidas, e, por essa razão, foram resultados de interações constantes com o meio social, podendo variar em relação a diferentes contextos em que foram construídas. Em termos gerais, a autora caracteriza as classificações das ciências segundo sete aspectos os quais foram sintetizados e serão brevemente apresentados nos tópicos seguintes. Segundo a autora, as classificações das ciências:

1. pressupõem um agente classificativo, que pode ser representado por um filósofo, tal como Bacon ou Peirce, por um cientista, como Ampère, por um elaborador de encicloplédias, como Diderot e D'Alembert, entre outros agentes;
2. estruturam-se por meio de um mecanismo classificador, estabelecido para executar as operações concernentes à classificação;
3. pressupõem a adoção de um princípio de classificação, como pode ser ilustrado por Comte ao estabelecer uma classificação baseada na ordem histórica de sua constituição e distinção progressiva;
4. cumprem um propósito específico, como pode ser visto nas classificações de Bacon, Piaget e Comte, em relação à orientação normativa da atividade científica que visaram proporcionar;
5. são estabelecidas mediante um conjunto de elementos finitos, logo, dizem respeito a ciências constituídas e àquelas em vias de constituição em determinada época, ou nela já conjecturável;
6. são construídas sobre contextos de classificações precedentes, ou seja, integram a história da classificação das ciências;
7. geram algum produto externo à atividade classificatória, tais como a constituição de determinada nomenclatura e sistemas diagramáticos das ciências.

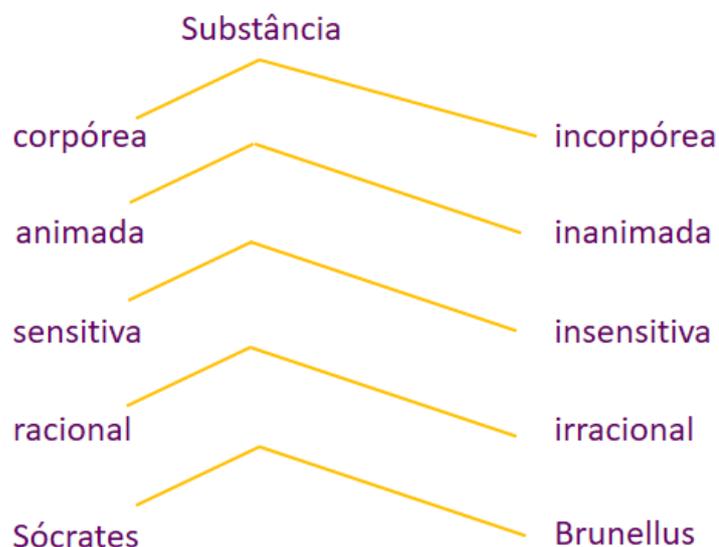
Tal como observado por Pombo (1998), Siqueira (2010) explica que a necessidade de sistematização do conhecimento, de modo a representar e ordenar o mundo, é antiga. Desde os períodos remotos, registros históricos demonstram o interesse da humanidade pela ordenação do conhecimento, a fim de dividi-lo ou classificá-lo. Nesse sentido, torna-se relevante destacar comentários tecidos por Mey e Silveira (2009) a respeito de biblioteca localizada em Ebla (atual Síria), que data do terceiro milênio antes de Cristo, descoberta em decorrência de missão arqueológica italiana na década de 1970. Em escavações realizadas foram encontradas entre 15.000 e 17.000 tábuas de argila que correspondiam a cerca de quatro mil documentos em escrita cuneiforme, devidamente ordenados em estantes de acordo com o conteúdo temático que apresentavam.

À semelhança do observado na biblioteca localizada em Ebla, Mey e Silveira (2009) também mencionam a biblioteca de Alexandria, que, por seu turno, reuniu cerca de 90.000 documentos dispostos entre aproximadamente 400.000 rolos de papiro, entre o provável período do século III a.C. ao século V d.C. Calímaco, erudito que trabalhava como bibliotecário em Alexandria, organizou o acervo dessa biblioteca em grandes assuntos, segundo a classificação aristotélica do conhecimento. De acordo com Campello (2006), Calímaco compilou um dos primeiros instrumentos de organização bibliográfica que se tem conhecimento – o *Pinakes*, que, por sua vez, era dividido por assuntos que abrangiam direito, literatura épica, medicina, matemática, ciências naturais, entre outros tópicos.

Ao se referir a classificações do conhecimento humano, Souza (2010) destaca que Platão teria sido o primeiro a agrupar este conhecimento de acordo com bases filosóficas. Aristóteles (384-322 a.C.), discípulo de Platão, defendia a ideia de que o conhecimento é vida e, sob essa perspectiva, classificou-o em teórico (Filosofia e Matemática), prático (Ciências Exatas e Sociais) e produtivo (Arte e Literatura).

Conforme assinalado por Pombo (1998), Porfírio partiu da trajetória classificatória estabelecida por Aristóteles e criou um esquema de classificação constituído por um conjunto hierárquico finito de gêneros e espécies, nos quais observam-se sucessivas dicotomias, como ilustrado na Figura 1:

Figura 1 – Esquema ilustrativo da árvore de Porfírio



Fonte: Adaptada de Pombo (1998, p. 7)

Tal como apresentado na Figura 1, Pombo (1998) explica que o esquema dicotômico proposto por Porfírio opera do nível geral ao específico, ou seja, do gênero à espécie. Nesse sentido, trata-se de uma classificação cuja influência se fez sentir em períodos posteriores à Idade Média, principalmente no que se refere à taxonomia dos seres resgatada pela ordenação evolutiva e à elaboração de classificações das ciências estruturadas sobre conceitos dicotômicos.

Segundo Souza (2010), Francis Bacon (1561-1626) divergia do ensino da filosofia aristotélica, publicando vários livros os quais se destaca o título *Novum Organum*, em contraste ao título *Organum*, que compila tratados sobre lógica elaborados por Aristóteles. Em síntese, Bacon classificou as ciências baseando-se nas faculdades humanas da memória, imaginação e razão.

Siqueira (2010) explica que o mundo ocidental passou por profundas mudanças entre os séculos XV e XVI, as quais destacam-se o desenvolvimento das cidades, das atividades comerciais e manufatureiras, além da disseminação da cultura humanista principalmente nas ciências e nas artes. Por conseguinte, novas áreas apontaram no âmbito do conhecimento científico, as quais contribuíram para o surgimento de diferentes formas de classificação.

No início da Modernidade, os estudos classificatórios foram fortemente influenciados por taxonomias, e, muitas vezes, divergiam entre si sob diversos aspectos. A título de exemplo, é possível remeter-se à corrente histórica em contraste com a prática, ao conhecimento de natureza pública em oposição ao privado, ao estabelecimento de superioridade de algumas ciências em relação a outras e ao conhecimento especializado em detrimento do conhecimento generalizado (SIQUEIRA, 2010).

No século XIX, ascende no contexto filosófico uma nova corrente teórica denominada Positivismo, criada por Augusto Comte (1798-1857). De acordo com Souza (2010), essa corrente classificou o conhecimento em três estágios: o teológico, o metafísico e o positivo. Além disso, estabeleceu o conceito hierárquico entre as ciências, no qual cada ciência que surge depende daquela que a precedeu. Ou seja, adotou o princípio da complexidade crescente e da generalidade decrescente. Todavia, Bhattacharjee (2012) explica que no decorrer do século XX, o Positivismo foi alvo de

severas críticas formuladas, principalmente, por sociólogos interpretativistas cujas origens remetiam à escola de pensamento do idealismo alemão. As críticas referiam-se à excessiva adoção de métodos quantitativos em pesquisas, em detrimento da utilização de métodos qualitativos, uma vez que aspectos relativos a ações sociais são igualmente relevantes para o estudo de determinados fenômenos, sob o ponto de vista interpretativo.

Souza (2010) ainda destaca a classificação criada pelo filósofo alemão Rudolf Carnap (1891-1970), na qual as ciências foram divididas em formais e factuais. As primeiras referem-se ao estudo de entes ideais, existentes em nível conceitual e abstrato, como pode ser visto no campo da Filosofia e da Matemática. Por seu turno, as ciências factuais lançam mão de fatos e observações para a sua validade.

Considerando o surgimento de diferentes classificações das ciências ao longo do tempo, com base em estudos realizados (MORRIS, 1935; HEMPEL, 1965; GREGORY, 1942; HELMER; RESCHER, 1959; HEDGES, 1987; WINDELBAND, 1980), Wazlawick (2014) aborda algumas dessas classificações, as quais serão brevemente discutidas nos parágrafos seguintes. Sendo assim, irá se discutir a respeito da distinção entre ciências formais e empíricas, ciências puras e aplicadas, ciências exatas e inexatas, ciências *hard* e *soft*, e ciências nomotéticas e idiográficas.

Segundo Wazlawick (2014), a distinção entre ciências formais e empíricas é um dos critérios classificatórios das ciências mais conhecidos. Em termos gerais, as ciências formais concentram-se no entendimento das ideias, independentemente de sua aplicação prática. Logo, seu objeto de estudo está na forma, ou nos processos lógicos ou matemáticos que as abrangem. A Matemática, a Estatística e alguns campos da Linguística Ilustram essas ciências. Por seu turno, as ciências empíricas, também conhecidas por reais ou factuais, concentram seus esforços no estudo de objetos e situações reais, lançando mão de experimentos e observações para fundamentação de suas descobertas.

Além disso, as ciências empíricas podem ser divididas em dois grupos os quais estão incluídas as ciências naturais e as ciências sociais. As primeiras referem-se ao estudo da natureza, independentemente da existência ou ação humana, como pode ser visto no campo da Astronomia, Física e Ciências da Terra. As segundas estão

relacionadas com o estudo da vida social dos indivíduos, como pode ser visto na área da Sociologia, Antropologia e Comunicação Social (WAZLAWICK, 2014).

Outra classificação bastante comum entre as ciências é aquela que distingue as ciências puras, também conhecidas por fundamentais, das ciências aplicadas. As primeiras estão relacionadas ao estudo de conceitos básicos do conhecimento, sem a preocupação direta em aplicá-los. Ou seja, o sentido de puro refere-se ao interesse precípuo nos fundamentos, leis ou ideias, como pode ser claramente observado no campo da Lógica. As segundas, por outro lado, dedicam-se à realização de estudos cujas descobertas são aplicadas diretamente em algum tipo de processo, tais como os relacionados ao contexto industrial (WAZLAWICK, 2014).

Referente à classificação que diferencia as ciências exatas das inexatas, pode-se dizer que as primeiras se caracterizam, sobretudo, pelo estudo de leis bastante previsíveis, uma vez que experimentos são replicados e produzem os mesmos resultados, como pode ser visto no campo da Física e da Química. As segundas, por seu turno, podem até prever aspectos de determinados fenômenos, entretanto, os resultados obtidos nem sempre ocorrem como o esperado, tendo em vista a ocorrência de fatores incontrolláveis que influenciam tais resultados (WAZLAWICK, 2014).

Com relação ao rigor dos métodos utilizados, Wazlawick (2014) explica que as ciências se distinguem em *hard* e *soft*. De acordo com o autor, as primeiras caracterizam-se, sobretudo, pela aplicação de métodos rigorosos em suas observações, experimentos e deduções, tal como pode ser observado no campo das Ciências Naturais e da Medicina. Por outro lado, nas ciências consideradas *soft* é muito difícil e, em alguns casos, até mesmo impossível, realizar pesquisas cujo ambiente seja totalmente controlado.

Por fim, as ciências podem ser classificadas como sendo nomotéticas ou idiográficas. As primeiras referem-se ao estudo de fenômenos que se repetem, e, por conseguinte, podem levar a descobertas de leis gerais em relação a diversos fenômenos. As segundas, por seu turno, investigam fenômenos únicos, ou seja, que não se repetem. Tal característica pode ser observada no campo da História, no qual

fatos não são repetitivos e não há padrões suficientemente determinantes para a declaração de leis gerais (WAZLAWICK, 2014).

Tendo em vista as transformações socioeconômicas, políticas e culturais ocorridas no século XX, principalmente em decorrência da segunda guerra mundial (1914-1939), Siqueira (2010) relembra que as formas de organização das áreas do conhecimento também foram modificadas. Diante disso, o desenvolvimento tecnológico, as descobertas científicas no contexto contemporâneo e o grande volume de informações produzido são aspectos que deflagram a necessidade de novas classificações das ciências, considerando esta realidade pós-moderna.

Conforme expõe Siqueira (2010), as inovações decorrentes do século XXI trazem consigo inúmeras possibilidades para o surgimento de novas classificações das ciências, tendo em vista que o conhecimento não é estanque. Tais classificações serão delineadas a partir dos traços culturais, sociais e ideológicos que trazem em seu bojo, oferecendo diversos recortes de realidades, sem a pretensão de estabelecer verdades absolutas em relação à organização do conhecimento.

Diante das explicações nas quais foram abordados aspectos relevantes ao desenvolvimento histórico das áreas do conhecimento e das classificações da Ciência, a seção seguinte apresentará reflexões com base na literatura científica a respeito das diferenças disciplinares sob o ponto de vista da comunicação científica, que por sua vez, é tópico fundamental ao contexto deste trabalho.

2.1.1 Diferenças disciplinares sob o ponto de vista da comunicação científica

Segundo Japiassu (1976), o que se entende por disciplina e disciplinaridade nada mais é do que a progressiva exploração científica de determinada área ou domínio do conhecimento. À vista disso, uma disciplina deveria estabelecer as fronteiras que a constitui. Tais fronteiras determinam quais são os objetos estudados, os métodos e sistemas adotados, os conceitos e teorias aplicados. Nesse sentido, o autor expõe alguns critérios utilizados para caracterizar a natureza de uma disciplina científica, os quais serão destacados nos tópicos seguintes:

1. o domínio material das disciplinas, constituído pelo conjunto de objetos pelos quais elas se interessam e dos quais se ocupam;
2. o domínio de estudo que nada mais é senão o ângulo específico sob o qual a disciplina considera seu domínio material, podendo ser comum a várias disciplinas (corresponde àquilo que os escolásticos chamavam de “objeto normal”);
3. o nível de integração teórica dos conceitos fundamentais e unificadores de uma disciplina, capazes de abranger todos os fenômenos próprios, tendo em vista uma reconstrução da “realidade” do domínio de estudo a fim de explicar e prever os fenômenos que a ele se referem;
4. os métodos próprios para apreender e transformar os fenômenos, havendo perfeita concordância entre a aplicação dos métodos e as leis gerais do nível de integração teórica;
5. os instrumentos de análise que repousam, sobretudo, na estratégia lógica, nos raciocínios matemáticos e na construção de modelos;
6. as aplicações das disciplinas: quanto mais elas se orientam para a aplicação profissional, mais ecléticas se revelam em sua concepção epistemológica, exigindo, assim, programas pluridisciplinares;
7. as contingências históricas: em seu processo de evolução histórica, cada disciplina se encontra, em cada fase, num momento de transição, em contato com forças e influências internas e externas do “aqui” e do “agora” (JAPIASSU, 1976, p. 60-61, *grifo do autor*).

Knorr-Cetina (1999) enfatiza as diferenças disciplinares nos domínios do conhecimento, reconhecendo-as como divisões culturais na Ciência, as quais foram operacionalmente denominadas: “culturas epistêmicas”. Tais culturas dispõem em tela diferentes cenários de conhecimento que refletem as práticas de especialistas concernentes a padrões e dinâmicas adotados em suas pesquisas. Cabe ressaltar que as divisões culturais no campo científico se tornam evidentes somente quando se entende a diversidade dos processos de conhecimento em diferentes domínios.

Meadows (1999, p. 60) assinala que “há diferenças autênticas entre disciplinas em função do tipo de conhecimento que os pesquisadores buscam e como eles o tratam”. Por esse ângulo, o autor enfatiza a heterogeneidade de práticas em relação às maneiras pelas quais pesquisas são conduzidas e seus resultados comunicados, aos requisitos para a aceitação de artigos em periódicos científicos, ao uso de computadores por pesquisadores e à organização administrativa de faculdades e departamentos em universidades.

Conforme o entendimento de Costa (2008), padrões de comunicação distinguem-se entre si em razão de diferenças disciplinares. Sob essa perspectiva, disciplinas têm sido classificadas em divisões do conhecimento ao longo do tempo,

de modo a considerar variações que dizem respeito a aspectos históricos e geográficos.

Dentre os tópicos de discussão relevantes a esse contexto, Costa (2008) ressalta que disciplinas diferentes tendem a demandar abordagens distintas no âmbito de serviços de informação, principalmente em relação ao ambiente digital, no que se refere à formação e desenvolvimento de coleções e ao formato de documentos e conjunto de dados, como ilustrado em seu comentário:

Disciplinas diferentes tendem a requerer estratégias diferentes, especialmente ao se lidar com o meio digital, como, por exemplo, na configuração de coleções, no formato de documentos, nos conjuntos de metadados, dentre inúmeros outros aspectos. Desse modo, quem quer que tenha de planejar e operar qualquer serviço de informação, tem que levar em conta diferenças disciplinares como uma questão vital. Tanto em termos conceptuais quando em termos técnicos, as diferenças disciplinares devem guiar as ações concernentes a qualquer ferramenta, recurso ou instrumento para lidar com a comunicação da pesquisa ou a gestão do conhecimento. Elas estão, geralmente, relacionadas ao contexto em que se realiza tanto o trabalho de pesquisa quanto a comunicação de seus resultados (COSTA, 2008, p. 224, *grifo nosso*).

Semelhantemente, Leite (2012) destaca que o reconhecimento das diferenças disciplinares quanto aos modos de produção do conhecimento, acesso e comunicação da informação por pesquisadores torna-se componente imprescindível a sistemas de gestão da informação científica. Em pesquisa cujo objetivo foi a proposição de modelo genérico de gestão da informação científica para instituições de pesquisa, Leite (2012) selecionou pesquisadores do Centro Brasileiro de Pesquisas Físicas (CBPF) e do Museu de Astronomia e Ciências Afins (MAST) para participarem do estudo.

Dentre os resultados obtidos pelo autor, dois aspectos podem ser destacados no tocante às diferenças disciplinares. O primeiro é relativo ao fato de que para físicos, o artigo de periódico científico em suas versões digitais (*e-prints*) foi considerado o principal insumo de informação para a condução das atividades de pesquisa, enquanto para cientistas sociais e humanistas, artigos publicados em anais de conferências e livros são igualmente relevantes para esse fim. O segundo diz respeito à publicação de resultados de pesquisa, no qual se observou que físicos publicam tais resultados preferencialmente em artigos de periódicos científicos, enquanto cientistas sociais e humanistas o fazem equiparadamente em artigos de

periódicos científicos, artigos disponibilizados em anais de conferências, edições organizadas de capítulos de livros, e, com menor frequência, em livros (LEITE, 2012).

Mueller (2005) aponta as diferenças disciplinares como aspectos relevantes ao contexto da comunicação científica. A título de exemplo, a autora ressalta que pesquisas desempenhadas no campo das Ciências Normais ou Experimentais geralmente sustentam-se sobre paradigmas universalmente aceitos, são conduzidas por equipes com grande número de pesquisadores e seus resultados geram artigos cuja preferência de publicação ocorre em periódicos internacionais e de língua inglesa. Por outro lado, no campo das Ciências Sociais e das Humanidades, muitas pesquisas são de autoria única e seus resultados podem gerar livros, capítulos de livros, entre outros registros do conhecimento, os quais não são necessariamente artigos publicados em periódicos científicos. É interessante notar, ainda, que neste campo há coexistência de diversas teorias e linhas de pensamento, assim como a adoção de diferentes métodos de pesquisa.

Pacheco, Abreu e Bernardino (2018) conduziram estudo vinculado ao grupo de pesquisa em Discurso, Identidade e Letramento Acadêmicos (DILETA) com os seguintes objetivos: 1) discutir como a disposição sociorretórica da seção de referências é apresentada no campo disciplinar da Linguística, Nutrição e Psicologia, já caracterizada em pesquisas de Costa (2015), Pacheco (2016) e Abreu (2016), respectivamente; 2) examinar os *corpora* analisados por esses autores nas culturas disciplinares referidas e investigar, por meio da seção de referências, as fontes de dados mais utilizadas e estimadas pelas áreas como fundamento de suas pesquisas. Para atingi-los, Pacheco, Abreu e Bernardino (2018) revisitaram e analisaram este *corpora*, constituído por 70 artigos acadêmicos publicados em periódicos indexados na plataforma *Webqualis* CAPES, entre 1999 e 2015, referentes às três disciplinas.

Dentre os principais resultados obtidos na pesquisa de Pacheco, Abreu e Bernardino (2018), podem ser destacados os seguintes tópicos:

- a citação de artigo acadêmico é predominante nas áreas mencionadas, sobretudo na área de Nutrição, correspondendo a quase 70% do total das citações identificadas. Isso reforça a valorização desse gênero da literatura científica pela referida área;

- a utilização do livro como fonte de informação em pesquisas demonstrou-se presente nas três áreas mencionadas, principalmente, na Linguística e Psicologia. É interessante notar que, geralmente, artigos no campo da Linguística destinam grande parcela de seu texto à exposição de fundamentos teóricos procedentes de livros. No que tange ao campo da Psicologia, observa-se que o significativo número de citações a livros nos artigos analisados pode ser explicado em razão de algumas subáreas, como a Psicanálise, terem o livro como fonte de informação de maior impacto, em detrimento de outros recursos. Com relação aos artigos analisados da área de Nutrição, foi percebido que esses não citam livros com frequência. Isso devido, principalmente, ao fato de o livro requerer muito tempo para ser publicado, não contendo, muitas vezes, os resultados de pesquisa mais recentes da área;
- especificamente em relação à área de Psicologia, observou-se a citação (ainda que pouco expressiva) de outras fontes de informação além de artigos acadêmicos, livros, teses e dissertações, dentre as quais podem-se destacar manuais, relatórios técnicos e anais de congresso.

Siemens, Liu e Smith (2014) exploraram em seus estudos o efeito das diferenças disciplinares sobre a colaboração científica. Com base em revisão de literatura realizada, os autores propuseram uma estrutura conceitual que auxilia o desenvolvimento de espaços colaborativos, considerando a influência de fatores que poderão melhorar ou oferecer danos às relações de trabalho as quais se deseja. Em um espaço de colaboração, duas dimensões se destacam: as diferenças disciplinares e a equidade de controle acadêmico entre membros de um projeto de pesquisa.

Referindo-se especificamente à dimensão das diferenças disciplinares, Siemens, Liu e Smith (2014) explicam que em um projeto científico colaborativo desenvolvido por pesquisadores de uma mesma disciplina, a linguagem e os procedimentos metodológicos são comuns a todos da área. Conseqüentemente, menores serão o tempo e os esforços consumidos na coordenação da pesquisa. Entretanto, a homogeneidade pode inibir a diversidade entre pesquisadores e, dessa forma, limitar potencialmente a criatividade e a inovação (SIEMENS; LIU; SMITH, 2014).

Logo, se um grupo de pesquisa valoriza tal pluralismo, é necessário que invista parte de seus recursos para o estabelecimento de um consenso em relação à linguagem utilizada, aos procedimentos metodológicos adotados e aos pontos de vista de um projeto de pesquisa, além da pacificação de possíveis pontos conflituosos (SIEMENS; LIU; SMITH, 2014).

Conforme o exposto, percebe-se que as diferenças disciplinares remetem a especificidades inerentes às áreas do conhecimento, que, por seu turno, dizem respeito à própria natureza das áreas, às suas tradições em termos de tipos de dados produzidos e às abordagens metodológicas adotadas. Por essa razão, a seguinte seção discorrerá sobre as diferenças disciplinares no que se refere aos dados de pesquisa.

2.1.2 Diferenças disciplinares em relação aos dados de pesquisa

Estudos como o de Borgman (2015) e o de Sayão, juntamente com Sales (2015), demonstram que não há uma definição ampla o suficiente para abranger várias perspectivas nas quais os dados de pesquisa podem ser entendidos. São inúmeras as contribuições que esses insumos informacionais podem oferecer à prática da pesquisa científica em diversas áreas do conhecimento, tais como a possibilidade de validação de resultados obtidos, replicação de pesquisas e a combinação a outros dados. Contudo, é cada vez mais evidente o fato de que dados são melhor compreendidos à luz das próprias culturas disciplinares nas quais foram gerados.

Com base nos estudos referidos, observa-se que os dados de pesquisa se diversificam quanto aos tipos, formatos, modos de produção, requisitos para armazenamento, entre outros aspectos, principalmente em razão da natureza disciplinar a qual pertencem e das características dos fenômenos estudados em cada área. Por isso, os dados podem incluir e não se restringem a: anotações de trabalho em campo; cadernos de laboratório; imagens; textos; algarismos; *softwares*; objetos; fósseis; artefatos; fotografias; vídeos; registros de questionários e entrevistas; equações e simulações.

Como forma de ilustrar a influência das diferenças disciplinares sobre processos inerentes à comunicação científica no que se refere aos dados de pesquisa, é interessante recorrer aos principais resultados de pesquisa publicada por Carvalho e Leite (2019), cujo objetivo foi identificar as práticas de produção/obtenção, compartilhamento e (re)uso de dados de pesquisa em Química, Antropologia e Educação. Para tanto, o estudo lançou mão do método de levantamento, e dados foram coletados por meio da realização de entrevistas semiestruturadas conduzidas com docentes permanentes dos programas de pós-graduação de Química (três pesquisadores), Antropologia (três pesquisadores) e Educação (quatro pesquisadores) da Universidade de Brasília (UnB), no mês de dezembro de 2017.

Com relação à produção de dados de pesquisa, observou-se entre os pesquisadores de Química entrevistados que os dados são fundamentalmente numéricos, incluindo modelos matemáticos e sistematização de compostos químicos. Já entre os pesquisadores de Antropologia, verificou-se que os dados de pesquisa são gerados por meio de diferentes procedimentos metodológicos, como observação participante, investigação bibliográfica, realização de entrevistas e redação de diários de campo. Consequentemente, há diversidade nos dados produzidos neste domínio disciplinar, incluindo dados qualitativos e quantitativos. Todavia, verificou-se o predomínio dos primeiros em detrimento dos segundos. Semelhantemente, a diversidade mencionada também foi notada no campo da Educação, uma vez que os dados são constituídos por discursos, investigações documentais, decorrentes de *surveys*, entre outros (CARVALHO; LEITE, 2019).

No que se refere ao compartilhamento de dados de pesquisa, verificou-se que pesquisadores entrevistados nos três domínios disciplinares reputam os dados que produzem como sendo úteis a outros pesquisadores. Todavia, observou-se que a referida prática se circunscreve ao contexto dos próprios grupos de pesquisa, de professores e estudantes, e de pesquisadores que realizam estudos com temáticas afins. Dessa forma, pesquisadores de Química entrevistados destacaram as especificidades dos dados que coletam em suas pesquisas, o que contribui para que muitos desses não sejam pertinentes a outras disciplinas. De outra forma, pesquisadores da Antropologia ressaltaram que o compartilhamento de dados deve

ocorrer mediante a observância de aspectos éticos. Já os pesquisadores na área de Educação, enfatizaram a necessidade da criação de repositórios como um meio que viabiliza o compartilhamento de dados de pesquisa de maneira estruturada (CARVALHO; LEITE, 2019).

Acerca do reuso de dados de pesquisa, notou-se que, em geral, essa prática não é comum entre pesquisadores da área da Química. Com relação ao campo da Antropologia, observou-se que o reuso de dados ocorre predominantemente entre pesquisadores de linhas correlatas. Na Educação, tal prática é bastante útil para a realização de estudos longitudinais e comparativos, ocorrendo, precipuamente, entre integrantes de um mesmo grupo de pesquisa. Além disso, pesquisadores nesse domínio disciplinar interessados em reutilizar dados, atentam-se à sobrevida dos dados para não incorrerem em desatualização. Nessa direção, alguns dados dizem respeito a um período específico e representam o recorte de uma realidade singular que, muitas vezes, não pode ser transferida a circunstâncias temporais mais recentes (CARVALHO; LEITE, 2019).

Considerando o exposto, importante também se mostra o estudo de Sales e Sayão (2019) acerca das diferenças disciplinares na gestão de dados de pesquisa no contexto da denominada *Big science* (grande ciência, traduzida para o português) e *Small science* (pequena ciência, traduzida para o português). O objetivo da pesquisa realizada foi analisar as diferenças mais significativas relativas à homogeneidade, gestão, curadoria, compartilhamento e institucionalização dos dados de pesquisa produzidos pela grande e pequena ciência. Porém, antes é fundamental caracterizar e distinguir cada um dos dois segmentos científicos mencionados, respectivamente.

Segundo Sales e Sayão (2019), alguns domínios disciplinares exemplificam nitidamente a dinâmica ocorrida na grande ciência, tais como Astronomia, Física, Genômica e outras áreas no campo das Humanidades Digitais. Essas disciplinas caracterizam-se, sobretudo, pela utilização de instrumentos de pesquisa imponentes, vastos recursos para o desenvolvimento de estudos, longa duração de seus projetos, envolvimento de grandes equipes de pesquisadores, moderna estrutura de laboratórios e pelo *know-how* nas áreas do conhecimento que atuam. A grande ciência produz e faz uso de acentuadas quantidades de dados.

Por outro lado, a pequena ciência é reconhecida por sua autonomia, interdisciplinaridade e vínculo com setores produtivos da sociedade. Contudo, parcela considerável dos recursos informacionais produzidos nesse segmento científico não está disponível para acesso, em razão de geralmente permanecerem arquivados nos computadores pessoais de pesquisadores, sem nenhuma intervenção de gestão para oferecer-lhes visibilidade. Ao contrário de domínios disciplinares da grande ciência, em que a disseminação de dados padronizadas é uma realidade, o compartilhamento de dados no âmbito da pequena ciência não é uma prática recorrente e sofre a interveniência de fatores, como o interesse de pesquisadores e a disponibilidade de infraestruturas apropriadas (SALES; SAYÃO, 2019)

Sales e Sayão (2019) destacam que agências de financiamento à pesquisa, formuladores de políticas, instituições de pesquisa, entre outras entidades, têm concentrado seus esforços na produção intensiva de dados. Conseqüentemente, nota-se um desequilíbrio no qual grandes quantidades de pequenas coleções de dados produzidos em universidades e centros de pesquisa passam despercebidas no contexto da comunicação científica. Sobre a problemática apresentada em relação à gestão de dados de pesquisa no âmbito da grande e da pequena ciência, os autores comentam:

O problema que se apresenta é que a gestão de dados é geralmente colocada de forma uniforme para toda a ciência e que os perfis que caracterizam a pequena e a grande ciência são geralmente desconsiderados ou minimizados no planejamento e no desenvolvimento das e-infraestruturas de pesquisa. Existem diferenças contundentes que se iniciam nas formas de geração dos dados até o seu arquivamento, na preservação e nas formas de compartilhamento. As metodologias, instrumentos, fluxos de pesquisa, homogeneidade dos dados gerados, volume de investimentos e reconhecimento e recompensa aos pesquisadores condicionam as formas de visibilidade/invisibilidade dos dados gerados pela pequena e grande ciência (SALES; SAYÃO, p. 153, 2019).

Portanto, a diversidade de dados e dos padrões de produção, compartilhamento, uso e gestão adotados em relação a dados evidenciam dois extremos na cadeia produtiva da ciência. Em uma ponta, se percebe o aumento descomunal da geração e disseminação de dados homogêneos e padronizados, o que compõem um cenário desafiador em relação ao controle dos processos

executados na pesquisa científica. Na outra ponta, observa-se uma extensa cauda constituída em sua essência pela diversidade de dados, demonstrada em termos de estrutura, formatos, complexidade e domínio disciplinar. Tal heterogeneidade acarreta a necessidade de gestão e curadoria de dados, de maneira que essas práticas abranjam as especificidades culturais, sociais e tecnológicas características de cada área do conhecimento (SALES, SAYÃO, 2019).

Com o propósito de ilustrar a diversidade inerente ao contexto dos dados de pesquisa, as próximas seções abordarão aspectos relacionados a duas áreas do conhecimento diferentes: Ciências Sociais e Química, respectivamente.

2.1.2.1 Dados de pesquisa no campo das Ciências Sociais

Segundo informações disponibilizadas no *site* do *Consortium of European Social Science Data Archives (CESSDA)*², dados quantitativos ou qualitativos são produzidos no contexto da pesquisa em ciências sociais. Tais dados abrangem diferentes materiais, incluindo conjuntos de dados numéricos, relatórios de entrevistas, anotações de campo, documentos coletados em estudos etnográficos, entre outros.

Assim também, com frequência, dados são obtidos a partir de, ou sobre participantes humanos. Consequentemente, muitos dados pessoais merecem atenção especial por se tratarem de aspectos sensíveis. Em geral, dados pessoais sensíveis podem oferecer riscos aos direitos e liberdades fundamentais de indivíduos. Tais dados incluem, mas não se restringem à: origem racial ou étnica; opinião política; crença religiosa; filiação sindical; orientação sexual; dados genéticos; condição de saúde; entre outros (CESSDA).

Considerando o exposto, observa-se que os dados de pesquisa na área das Ciências Sociais caracterizam-se, sobretudo, por sua diversidade evidenciada pelos modos pelos quais são gerados e pelos contextos culturais os quais estão inseridos. Por conseguinte, é altamente relevante que esse aspecto seja reconhecido para a

2 Disponível em: <https://www.cessda.eu/Training/Training-Resources/Library/Data-Management-Expert-Guide/1.-Plan/Data-in-the-social-sciences>

devida gestão de dados no âmbito dessas ciências. Um caminho para isso é a realização de estudos que abordem práticas de pesquisadores em relação aos dados que produzem, como a pesquisa conduzida por Saeed e Ali (2019).

O estudo de Saeed e Ali (2019) objetivou analisar, avaliar e comparar as práticas de gestão e compartilhamento de dados de pesquisa entre pesquisadores da Faculdade de Ciências da Vida e Ciências Sociais da *Aligarh Muslim University* (AMU). Para atingi-lo, a pesquisa lançou mão do método de levantamento e aplicou questionários a 352 bolsistas de pesquisa dessas duas grandes áreas do conhecimento – 150 bolsistas no campo das Ciências da vida; 202 bolsistas no campo das Ciências Sociais. Os principais resultados obtidos em relação às Ciências Sociais podem ser apresentados da seguinte maneira:

- quanto aos tipos de dados produzidos, foi observado que a maior parcela das respostas obtidas apontaram que pesquisadores geravam dados provenientes de questionários (61,39%), estatísticas (60,4%) e decorrentes de respostas de *survey*;
- com relação ao Plano de gestão de dados, 72,28% dos pesquisadores assinalaram que possuíam um Plano de gestão de dados, entretanto, 25,74% apontaram que não faziam uso do plano;
- no que se refere ao armazenamento de dados, verificou-se que 85,15% do total das respostas obtidas assinalaram fazê-lo em dispositivos pessoais. A segunda resposta com o maior número de registros logrados (48,51%) corresponde à cópia impressa de dados. A porcentagem relativa ao armazenamento de dados em nuvem foi de somente 4,95%;
- no que tange à prática de compartilhamento de dados, 55,44% dos pesquisadores sinalizaram que compartilhavam os dados que produziam com terceiros;
- além disso, observou-se a utilização de diferentes métodos no compartilhamento de dados – 30,69% das respostas apontaram que pesquisadores compartilhavam seus dados em periódicos científicos, 29,7% por meio de redes sociais acadêmicas e 18,81% os disponibilizavam a seus pares quando solicitados. O depósito de dados em repositórios de dados abertos e

em um serviço de repositório institucional de dados de pesquisa foram as estratégias menos utilizadas para o compartilhamento de dados (cada uma delas correspondendo a 5,94% das respostas);

- acerca dos desafios enfrentados por pesquisadores em relação ao compartilhamento de dados de pesquisa, alguns inconvenientes foram os mais citados dentre as respostas obtidas, sendo eles: a dificuldade na preservação da privacidade e a confidencialidade de dados (37,62%); bem como o tempo e esforço necessários ao compartilhamento de dados (19,8%). Já questões técnicas relacionadas aos dados, foram mencionadas em apenas 3,96% das respostas obtidas.

2.1.2.2 Dados de pesquisa no campo da Química

Com a finalidade de caracterizar dados de pesquisa produzidos na Química e apresentar necessidades da área em relação à gestão de dados, Chen e Wu (2017) conduziram estudo a partir de cinco perspectivas: geração de dados; processamento; preservação; compartilhamento e necessidades relacionadas a gestão de dados. Para tanto, os autores aplicaram um questionário a 119 pesquisadores e estudantes de pós-graduação do Instituto de Química do *Chinese Academy of Science* (CAS), entre 2015 e 2016. Os dados obtidos foram analisados estatisticamente.

Dentre os resultados encontrados, Chen e Wu (2017) observaram que os respondentes geram diversos tipos de dados, sendo os dados experimentais os mais comuns, correspondendo a quase 80% do total de respostas. Em seguida, computaram-se os dados de observações (36,13%). Quanto aos formatos de dados, verificou-se que o mais adotado inclui espectogramas (72,27%), seguido de dados de testes de instrumentos experimentais (71,43%), imagens (59,66%), tabelas (47,90%), dados de processamento de software (46,22%) e textos (45,38%).

Chen e Wu (2017) também constataram que os participantes da pesquisa assinalaram que o caderno de laboratório impresso é o meio mais comum adotado para o registro de dados, correspondendo a 81,51% do total de respostas. Com base no resultado obtido, os autores comentaram que o caderno de laboratório eletrônico

e sistemas de gestão de dados de laboratório ainda não são meios de registros populares no contexto de laboratório. Além disso, os autores alertaram que os mecanismos tradicionais adotados pelos respondentes para o registro de suas pesquisas não contribuem para a preservação de dados a longo prazo, tampouco para o compartilhamento desses.

No que se refere à preservação de dados, Chen e Wu (2017) destacaram que 81,51% dos respondentes utilizam computadores pessoais para esse fim, seguido de cadernos de laboratório impressos (74,79%), de discos de armazenamento de dados (52,10%) e de computadores coletivos da equipe de pesquisa (47,06%). Os autores destacaram que apesar da comodidade de salvar dados em computadores e discos, há a desvantagem da vulnerabilidade de perda de dados em razão da obsolescência de equipamentos tecnológicos. Constatou-se, ainda, que apenas 2,52% dos respondentes depositaram dados em repositórios visando o armazenamento seguro desses dados.

No que se refere ao compartilhamento de dados quando pesquisadores são requisitados por periódicos científicos na área da Química, Chen e Wu (2017) observaram que os respondentes da pesquisa demonstraram as seguintes opiniões em relação a isso: 52,10% afirmaram estar extremamente dispostos a compartilhar seus dados, 38,66% estão dispostos a realizá-lo e 6,72% ainda não têm certeza.

De modo semelhante, Abduldayan *et al.* (2021) conduziram estudo com o objetivo de compreender as práticas de gestão de dados de pesquisadores da Química em cinco universidades de tecnologia federais na Nigéria. Para alcançá-lo, os autores realizaram entrevistas semiestruturadas com 40 pesquisadores das seguintes universidades: *Abubakar Tafawa Balewa University of Technology*, *Federal University of Technology - Akure*; *Federal University of Technology - Minna*; *Federal University of Technology - Owerri*; e *Modibbo Adama University of Technology - Yola*.

Dentre os resultados obtidos no estudo de Abduldayan *et al.* (2021), podem-se destacar:

- pesquisadores produzem e utilizam com frequência dados experimentais;

- houve relatos de várias experiências de perda irreparável de dados decorrente de armazenamento inadequado desses, ou não realização de cópias de segurança;
- pesquisadores não têm ampla conscientização sobre gestão de dados de pesquisa;
- a biblioteca foi mencionada como um local apropriado para o armazenamento seguro de dados a longo prazo, sendo também relevante na disponibilização de treinamentos e orientações gerais sobre o tratamento de dados;
- pesquisadores estavam receosos de que dados que produzem sejam acessados por terceiros não autorizados.

A partir dos resultados mencionados, Abduldayan *et al.* (2021) recomendaram que as bibliotecas das universidades nas quais atuam os participantes da pesquisa promovam discussões sobre a gestão de dados de pesquisa. Eventos como *workshops* e seminários podem evidenciar a importância e os benefícios desse tipo de gestão nas instituições. Ademais, as universidades poderiam elaborar políticas que estimulassem a submissão de dados a repositórios institucionais.

Além dos aspectos até então tratados, este trabalho pauta-se sobre a perspectiva da Ciência aberta, na medida em que propõe diretrizes para a gestão de dados abertos em diferentes áreas. Considerando tal perspectiva, torna-se necessário lançar mão de breve contextualização da Ciência aberta na seção seguinte, que cumprirá o propósito de apresentar importantes conceitos e definições relacionados à discussão desse tópico na literatura científica e técnica.

2.2 Breve contextualização da Ciência aberta

Considerando que o mundo tem enfrentado desafios complexos, como escassez de recursos naturais, alterações climáticas, pobreza, doenças e poluição ambiental, Cribb e Hartomo (2010) deflagram a necessidade do conhecimento humano ser compartilhado. Os autores revelam uma grande lacuna entre a produção do conhecimento científico e o seu compartilhamento, na medida em que parte significativa do esforço intelectual, do dinheiro e tempo empregados em pesquisas é

subutilizada, ou não atinge seu potencial. Assim, o conhecimento que poderia ser útil na resolução de desafios complexos vivenciados na atualidade, muitas vezes, encontra-se indisponível, com restrições de acesso, ou disposto em linguagem inacessível.

Em razão do exposto, Cribb e Hartomo (2010) defendem uma ciência mais aberta e democrática. Por isso, uma mudança cultural torna-se necessária nesse contexto para que o conhecimento leigo e o científico sejam igualmente relevantes nos processos de inovação do conhecimento. Além disso, ao utilizarem a democratização da Ciência como pano de fundo para compreender aspectos da Ciência aberta, os autores entendem que a sociedade pode trazer contribuições nesse sentido, tornando-se uma parceira importante na construção do conhecimento científico.

Cribb e Hartomo (2010) também sugerem iniciativas que podem colaborar com a Ciência aberta, as quais se destacam: a comunicação de pesquisas entre cientistas e a sociedade; o estabelecimento de sistemas de recompensa e o reconhecimento de práticas relacionadas à Ciência aberta; o comprometimento de pesquisadores com políticas de transparência pública e de Ciência aberta; a disponibilidade de recursos tecnológicos que otimizem a comunicação de pesquisas científicas no âmbito do acesso aberto.

Semelhantemente ao entendimento de Cribb e Hartomo (2010), Vicent-Saez e Martinez-Fuentes (2018) afirmam que a Ciência aberta está surgindo como um fenômeno disruptivo global, cujos efeitos são bastante significativos no contexto europeu. Sob essa perspectiva, os autores explicam que tal fenômeno acarretou mudanças socioculturais e tecnológicas baseadas no princípio da abertura e conectividade em relação à maneira pela qual a pesquisa é planejada, conduzida e avaliada. Nessa direção, ferramentas abertas de dados, plataformas abertas de acesso, métodos de revisão aberta por pares e atividade de interação com a comunidade não acadêmica são fatores que influenciam o contexto científico e podem acelerar o ciclo da pesquisa.

A partir do cenário delineado, ampliam-se as discussões sobre a Ciência aberta, e, por conseguinte, destaca-se a necessidade de contextualizar esse termo a partir de outros pontos de vista localizados na literatura científica e técnica. À vista disso,

no relatório intitulado: *Study on open science: impact, implications and policy options*, publicado em 2015 pelo *Directorate-General for Research and Innovation* da *European Commission*, a Ciência aberta pode ser entendida como uma nova abordagem no contexto científico alicerçada no trabalho cooperativo e no compartilhamento de informações por meio de redes colaborativas. Em síntese, a Ciência aberta estimula a criação de soluções com base na abertura e no compartilhamento de conhecimentos, em oposição à apropriação restrita de descobertas e avanços científicos.

O projeto *Open Responsible Research and Innovation to further Outstanding Knowledge* (acrônimo utilizado para Orium)³, financiado pelo programa de pesquisa e inovação da *European Union's Horizon 2020*, sob o programa de trabalho *Science With And For Society* (SWAFS), explora maneiras para que organizações de pesquisa em ciências da vida e biomedicina adotem práticas relacionadas à Ciência aberta. Sob esse ponto de vista, o projeto desencadeia mudanças institucionais, culturais e comportamentais.

De acordo com o projeto Orium⁴, a Ciência aberta é um movimento cujo foco está na disponibilização de pesquisas científicas e de seus dados a membros da comunidade acadêmica, a profissionais da informação e a cidadãos. Sustenta-se na expansão do acesso ao conhecimento científico, de modo a incluir práticas de acesso aberto ao longo do ciclo da pesquisa e a promover a colaboração entre pesquisadores. Além disso, a Ciência aberta pode contribuir para a qualidade e transparência de pesquisas, apropriação pública da ciência, e compartilhamento de dados de pesquisa e de recursos.

Informações retiradas do *site* do projeto Orion apontam cinco caminhos rumo à Ciência aberta: 1) acesso aberto às publicações; 2) dados abertos de pesquisa; 3) pesquisa aberta; 4) política de ciência aberta; 5) trabalho em comunidade. Esses caminhos são detalhados da seguinte maneira:

- 1) acesso aberto às publicações** – diferentemente do atual modelo de publicação científica, algumas instituições de pesquisa sem fins lucrativos e instituições

3 Disponível em: <https://www.orion-openscience.eu/about>

4 Disponível em: <https://www.orion-openscience.eu/resources/open-science>

educacionais estão adotando o paradigma do acesso aberto. É o que pode ser observado na disponibilização de periódicos em acesso aberto realizada pela *Public Library of Science* (PLOS) e pelo conjunto de artigos, sem viés editorial, disponibilizados pela plataforma *F1000Research*. Acrescente-se a esse contexto, oportunamente, o modelo cooperativo de publicação eletrônica de periódicos científicos *Scientific Electronic Library Online* (SciELO)⁵, produto de colaboração entre a Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), o Centro Latino-Americano e do Caribe de Informação em Ciências da Saúde (conhecido como BIREME), outras instituições relacionadas à comunicação científica e editores científicos. Nessa direção, o modelo SciELO proporciona visibilidade e amplo acesso à publicação científica produzida em países em desenvolvimento, especialmente, àquela gerada na América Latina e Caribe;

- 2) **dados abertos de pesquisa** – o acesso aberto aos dados de pesquisa é um dos passos principais para tornar a ciência aberta e transparente, de modo que o avanço ao conhecimento está intimamente relacionado à acessibilidade de dados;
- 3) **pesquisa aberta** – esse tipo de abordagem contribui para a reprodutibilidade de pesquisas, uma vez que a replicação de estudos permanece uma prática desafiadora no contexto científico. Outrossim, é possível aumentar a qualidade e o impacto de pesquisas, tendo em vista que o compartilhamento de trabalhos proporciona a avaliação aberta por terceiros, e, conseqüentemente, atribui maior qualidade, visibilidade e transparência à produção científica;
- 4) **política de ciência aberta** – é fundamental a proposição de políticas de ciência aberta claramente definidas;
- 5) **trabalho em comunidade** – considerando que a ciência investiga questões complexas, é importante que o pesquisador integre comunidades e organizações científicas para apreender uma visão compartilhada do processo de pesquisa. É interessante notar que o princípio do altruísmo é fundamental

5

Disponível em: https://wp.scielo.org/wp-content/uploads/Modelo_SciELO.pdf

para que comunidades científicas avancem em suas descobertas. Isso quer dizer que comunidades precisam reconhecer a interdependência entre seus membros, além de fomentar a empatia e a diversidade de interesses inerentes a diferentes áreas do conhecimento.

Segundo Jomier (2017), a Ciência aberta busca acelerar o processo de publicação científica, de maneira que o maior número de informações seja disseminado. Para tanto, a autora considera que a Ciência Aberta engloba três componentes principais: acesso aberto; dados abertos; código aberto; conforme ilustrado na Figura 2:

Figura 2 – Componentes da Ciência aberta



Fonte: traduzida e adaptada de Jomier (2017, p. 362)

Os três componentes da Ciência aberta ilustrados na Figura 2 são detalhados da seguinte maneira:

- **acesso aberto** – publicações em acesso aberto são alternativas à restrição de acesso adotada por alguns editores, de modo que a disseminação de pesquisas ocorra livremente. Apesar disso, ainda se observam periódicos de acesso aberto

que disponibilizam gratuitamente artigos a seus leitores somente após período de embargo estabelecido. Conseqüentemente, para ter acesso a artigos mais recentes, é necessário adquiri-los por meio de compra;

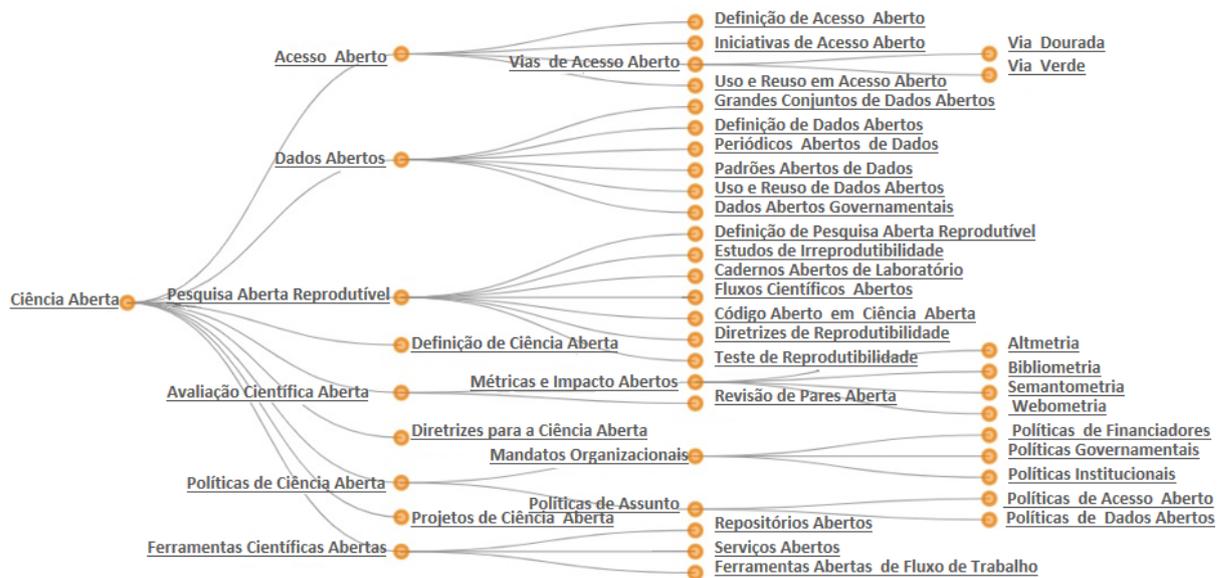
- **dados abertos** – referem-se ao compartilhamento de dados de pesquisa livre e abertamente. Isso porque já é conhecido que tais dados podem contribuir para a reprodutibilidade de pesquisas e para a redução significativa de tempo empregado em determinados experimentos. Todavia, alguns desafios destacam-se no contexto do compartilhamento de dados de pesquisa, como a dificuldade de pesquisadores em lidar com dados de diversas naturezas e oriundos de áreas de conhecimento distintas, a necessidade de aprimoramento de infraestruturas tecnológicas suficientemente robustas para armazenar grandes quantidades de dados e o estabelecimento de licenças de compartilhamento de dados de pesquisa;
- **código aberto** – o movimento de Código aberto caracteriza-se, sobretudo, por sete valores: segurança; acessibilidade; transparência; perpetuidade; interoperabilidade; flexibilidade; localização. Apesar de variações quanto ao escopo e à extensão de licenças, em geral, o contexto universitário e o industrial têm maximizado o uso de licenças de *software* de código aberto.

O projeto *Foster Plus*⁶ (acrônimo usado para *Facilitate Open Science Training for European Research*) foi desenvolvido por onze parceiros em seis países e financiado pela União Europeia com o propósito de contribuir para a transformação do comportamento de pesquisadores em relação à Ciência aberta. Para tanto, visa alcançar três objetivos: 1) apoiar uma mudança cultural em favor da Ciência aberta no contexto da pesquisa científica; 2) consolidar e sustentar uma rede de treinamento sobre Ciência aberta; 3) fortalecer capacitações sobre Ciência aberta. Portanto, esse projeto compromete-se em produzir conteúdos para treinamento sobre Ciência aberta, de maneira que esses possam ser devidamente incorporados à rotina de pesquisadores.

6 Disponível em: <https://www.fosteropenscience.eu/about#theprevproject>

A Figura 3 apresenta a taxonomia para a Ciência aberta⁷ criada pelo Foster, a qual é constituída por nove elementos principais: 1) acesso aberto; 2) dados abertos; 3) pesquisa aberta reprodutível; 4) definição de Ciência aberta; 5) avaliação científica aberta; 6) diretrizes para a Ciência aberta; 7) políticas de Ciência aberta; 8) projetos de Ciência aberta; 9) ferramentas científicas abertas.

Figura 3 – Taxonomia da Ciência aberta



Fonte: traduzida e adaptada de Foster

Os nove elementos principais que constituem a taxonomia para a Ciência aberta proposta pelo Foster são descritos da seguinte forma:

- 1) **acesso aberto** – acesso *on-line* e gratuito a conteúdos científicos revisados por pares, em consonância com direitos autorais e licenças de uso;
- 2) **dados abertos** – dados acessíveis *on-line* e gratuitamente. Podem ser reutilizados e compartilhados mediante atribuição de autoria;
- 3) **pesquisa aberta reprodutível** – aplicação direta da Ciência aberta e livre acesso aos elementos que compõem experimentos, viabilizando a reprodutibilidade da pesquisa;

- 4) **definição de Ciência aberta** – prática baseada em projetos colaborativos de pesquisas, caracterizada sobretudo pelo livre acesso, compartilhamento e reuso de dados de pesquisa, de registros de laboratórios e de outros processos de pesquisa;
- 5) **avaliação científica aberta** – avaliação sobre resultados de pesquisas realizada não apenas por pares, mas também pela comunidade não acadêmica;
- 6) **diretrizes para a Ciência aberta** – conjunto de diretrizes que abrangem as melhores práticas em relação à Ciência aberta;
- 7) **políticas de Ciência aberta** – elaboração de políticas que englobam as melhores práticas em consonância com objetivos fundamentais da Ciência aberta;
- 8) **projetos de Ciência aberta** – criação de projetos relacionados à Ciência aberta;
- 9) **ferramentas científicas abertas** – desenvolvimento de ferramentas que auxiliem nas atividades relacionadas à Ciência aberta.

De acordo com Fecher e Friesike (2014), o termo Ciência aberta tornou-se foco de vívidas discussões na comunidade científica, remetendo-se à abertura do processo de criação do conhecimento e dos resultados de pesquisas, e a maior interação entre estudos científicos e a sociedade. À vista disso, interferências em um novo contexto científico que se estabelece tornam-se cada vez mais presentes no âmbito de diversos *stakeholders*, tais como pesquisadores de diferentes áreas, decisores políticos, programadores e operadores de plataformas, editores científicos e o público interessado.

Lançando mão de estudos sobre a Ciência aberta presentes na literatura científica, Fecher e Friesike (2014) identificaram cinco escolas de pensamento distintas, denominadas: Escola democrática; Escola pragmática; Escola de infraestrutura; Escola pública; Escola métrica. O Quadro 1 apresenta cada uma dessas escolas conforme seu pressuposto principal, grupos envolvidos, objetivo principal, ferramentas e métodos.

Quadro 1 - Cinco escolas de pensamento da Ciência aberta

ESCOLA DE PENSAMENTO	PRESSUPOSTO PRINCIPAL	GRUPOS ENVOLVIDOS	OBJETIVO PRINCIPAL	FERRAMENTAS E MÉTODOS
Democrática	O acesso ao conhecimento é desigualmente distribuído	Cientistas, políticos, cidadãos	Tornar o conhecimento livremente disponível a todos	Acesso aberto, direitos de propriedade intelectual, dados abertos, códigos abertos
Pragmática	A criação do conhecimento poderia ser mais eficiente se cientistas trabalhassem juntos	Cientistas	Abertura do processo de criação do conhecimento	Saber popular, efeitos das redes de relacionamento, dados abertos, códigos abertos
Infraestrutura	A pesquisa eficiente depende de ferramentas e aplicações disponíveis	Cientistas e provedores de plataformas	Criar plataformas abertamente disponíveis, ferramentas e serviços para cientistas	Plataformas colaborativas e ferramentas
Pública	A ciência precisa se tornar acessível ao público	Cientistas e cidadãos	Tornar a ciência acessível aos cidadãos	Ciência Cidadã, relações públicas da ciência, blogs científicos
Métrica	As contribuições científicas hoje necessitam de métricas de impacto alternativas	Cientistas e políticos	Desenvolver um sistema métrico alternativo para impacto científico	Altmetrias, revisão por pares, citação, fatores de impacto

Fonte: traduzido e adaptado de Fecher; Friesike (2014, p. 20)

O foco da Escola democrática está no livre acesso ao conhecimento, especialmente quando pesquisas forem financiadas por recursos estatais. Sob essa perspectiva, defende-se o acesso aberto a publicações e a dados de pesquisa. A Escola pragmática, por seu turno, argumenta a favor de tornar a pesquisa mais eficiente, na medida em que a ciência é entendida como um processo que pode ser otimizado. Para tanto, é importante a modularização do processo de criação do conhecimento, a inclusão de conhecimentos externos e a colaboração viabilizada por meio de ferramentas *on-line* (FECHER; FRIESIKE, 2014).

A Escola de infraestrutura defende a viabilização de infraestrutura técnica com a finalidade de intensificar práticas colaborativas no ambiente da Internet, por meio do desenvolvimento de *softwares*, ferramentas, aplicações e redes de computação. Sob esse ponto de vista, podem ser delineadas duas tendências denominadas: computação distribuída; redes sociais e colaborativas para cientistas. A primeira refere-se à utilização de recursos computacionais em projetos de pesquisas caracterizados

por intensa produção de dados, altas taxas de transferências de arquivos e necessidade de armazenamento. O foco da segunda está no pesquisador, de modo que plataformas lhe proporcione maior interação e colaboração com outros pesquisadores (FECHER; FRIESIKE, 2014).

A Escola pública deflagra a necessidade de a ciência tornar-se acessível a um público mais amplo. Pressupõe que tecnologias decorrentes da *Web 2.0* viabilizam a abertura do processo de pesquisa e a disseminação de resultados de pesquisas a público não-especialista. Além disso, a Escola pública pode ser categorizada em duas correntes de pensamento, as quais se referem à acessibilidade ao processo de pesquisa (a produção) e à compreensão de resultados da pesquisa (tornar a ciência entendível). A primeira está relacionada com o conceito de “ciência cidadã”, que presume a participação de público leigo e amadores na pesquisa científica, e com o financiamento colaborativo da ciência, no qual usuários da Internet contribuem com dinheiro para a realização de pesquisas de interesse social. A segunda, por seu turno, refere-se à importância de cientistas publicarem os resultados de suas pesquisas com uma linguagem compreensível à ampla audiência, ou seja, entendível ao público não especialista (FECHER; FRIESIKE, 2014).

A Escola métrica defende a adoção de padrões alternativos na avaliação do impacto científico, na medida em que o fator de impacto tem se tornado uma variável cada vez mais decisiva para o financiamento de pesquisas e oportunidades na carreira de pesquisadores. Sob esse ponto de vista, argumenta-se que a revisão por pares é dispendiosa em relação ao tempo demandado para tal atividade, o impacto é aferido preponderantemente ao periódico científico, e não ao artigo publicado, e que novos formatos de publicação raramente são avaliados no tocante ao impacto que geram. Em razão disso, defende-se a adoção de métricas de impacto alternativas e mais ágeis que incluam publicações não convencionais e abranjam contribuições científicas disponibilizadas em ambiente da *web* (FECHER; FRIESIKE, 2014).

Portanto, como se observa nos estudos abordados sobre a Ciência aberta apresentados até o momento, há diversos conceitos, definições e linhas interpretativas que contribuem para a discussão em nível teórico a respeito desse tópico. Além disso,

é notável a indissociabilidade entre Ciência aberta e acesso aberto. Pela característica mencionada, a próxima seção tratará, especificamente, sobre o acesso aberto.

2.2.1 Considerações acerca do acesso aberto

De acordo com Mueller (2006), a comunidade científica atribuiu às revistas revisadas por pares a posição de canais preferenciais para a validação do conhecimento produzido e confirmação de autoria de descobertas. Todavia, esse fato não foi suficiente para conter o que desestabilizou o tradicional sistema de comunicação científica em nível global, quando irrompeu a chamada crise dos periódicos em 1980, pré-anunciada na década anterior.

Diversos fatores culminaram na crise dos periódicos, os quais podem-se destacar as dificuldades de bibliotecas universitárias e centros de pesquisas (principalmente americanos) em manter atualizadas suas coleções de periódicos científicos, e ainda em suprir as necessidades de informação de seus usuários, além de impasses financeiros vivenciados por editoras, devido aos altos custos de produção de periódicos (MUELLER, 2006).

Ao se referir ao modelo de negócio que caracteriza o tradicional contexto da publicação periódica científica, Costa (2006) destaca que esse modelo consiste em cinco ações principais, sendo elencadas da seguinte maneira:

- autores recebem financiamento de suas instituições ou de agências de fomento para iniciar suas pesquisas;
- após a realização de pesquisas, autores redigem os resultados obtidos e remetem seu manuscrito a editores;
- solicitados por editores, especialistas nas áreas concedem parecer sobre o manuscrito recebido;
- caso o manuscrito seja aceito, esse é publicado pelos editores;
- o acesso ao trabalho do autor restringe-se ao custeio da assinatura do periódico em que a pesquisa foi publicada.

Diante do exposto, Costa (2006) tece comentários ao que a própria autora chamou de “perversidade do modelo”. Nesse sentido, a crítica que se estabelece

reside no fato de universidades e agências de fomento financiarem pesquisas científicas e os resultados publicados não serem livremente acessíveis a seus autores, unidades de informação, ou centro de pesquisas. Outro aspecto relevante a essa discussão concentra-se no princípio no qual pesquisas subsidiadas por recursos públicos devem ser amplamente acessíveis. Entretanto, o modelo de negócio atualmente assumido por editores comerciais atribui preços exorbitantes e impõem restrições de acesso a publicações de pesquisas, geralmente, financiadas por recursos públicos.

Contudo, Mueller (2006) destaca que, no fim da década de 1980, ganham força neste cenário discussões sobre alternativas às restrições ao acesso a publicações científicas, considerando as novas possibilidades tecnológicas vislumbradas à época. Desse modo, tais publicações em meio eletrônico trouxeram a esperança de uma transformação no tradicional sistema de comunicação científica. Sobre esse aspecto Mueller (2006, p. 27) tece o seguinte comentário:

Assim como os utopistas da Renascença, alguns sonharam com um novo sistema de comunicação, no qual o acesso a todo conhecimento científico se tornaria universal e sem barreiras. Especialmente nos países mais afastados dos principais centros produtores, surgiu a esperança não só de acesso ao que era produzido fora, mas também que a produção local teria maior visibilidade e penetração internacional (MUELLER, 2006, p. 27).

Os primeiros periódicos eletrônicos surgiram na década de 1990, período que também coincide com o aparecimento daqueles em acesso livre. Nesse momento, não havia mudanças significativas entre periódicos eletrônicos por assinatura e periódicos eletrônicos de acesso livre em relação à apresentação, estrutura e leiaute das publicações. No entanto, já havia uma diferença decisiva entre ambos, que era a possibilidade de pesquisadores acessarem artigos de periódicos científicos sem pagamento prévio (MUELLER, 2006, p. 27).

Para Alonso Arévalo, Subirats Coll e Martínez Conde (2008), a crise dos periódicos foi o gatilho para desencadear o movimento de acesso aberto como alternativa às barreiras de acesso impostas por editores. Entretanto, tal fenômeno só se tornou possível, mediante o desenvolvimento e uso de redes de informação que viabilizaram o compartilhamento imediato dos registros do conhecimento científico.

A década de 1990 também se caracteriza pelo surgimento de iniciativas isoladas de pesquisadores e gestores da informação, cujo propósito era promover o acesso aberto às publicações científicas. Dentre essas iniciativas, destaca-se a criação do arquivo aberto de *pré-prints* chamado: ArXiv, em 1991, pelo físico Ginsparg. Decorrendo a década, mudanças ocorreriam no modelo de comunicação científica e seriam responsáveis pela concretização do que hoje é conhecido por acesso aberto. (ALONSO ARÉVALO; SUBIRATS COLL; MARTÍNEZ CONDE, 2008).

Por sua vez, os anos 2000 tornaram-se palco de discussão sobre o acesso aberto, não somente aos resultados de pesquisas, mas também aos conjuntos de dados que as subsidiam ao longo de todo o processo da produção científica. Os parágrafos seguintes empenham-se em discutir, de modo lacônico, algumas iniciativas no contexto internacional e brasileiro que se destacam em relação à defesa do acesso aberto à literatura científica e aos dados de pesquisa. Dentre essas, serão abordados na seguinte sequência: a Iniciativa de Acesso Aberto de Budapeste (2002) (em inglês, *Budapest Open Access Initiative* – BOAI); a Declaração de Bethesda (2003); a Declaração de Berlim sobre o acesso livre ao conhecimento nas Ciências e Humanidades (2003); os Princípios *Panton* (2010); o Manifesto de acesso aberto a dados da pesquisa brasileira para ciência cidadã (2016), produzido pelo Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT) e o Manifesto bibliotecário pela ciência aberta na América Latina (2019).

Em 2001, o *Open Society Institute* (OSI) conduziu uma reunião em Budapeste com os principais atores do sistema da comunicação científica proponentes ao acesso aberto à literatura acadêmica e científica. Dentre esses, foram convocadas instituições governamentais, universidades, bibliotecas, editores, pesquisadores e sociedades científicas. Como resultado da reunião, um documento foi publicado em 2002, o qual demonstrou a relevância da discussão sobre a disponibilidade *on-line* gratuita da informação científica a todos, de maneira irrestrita (BUDAPEST OPEN ACCESS, 2002).

Por acesso aberto à literatura científica, entende-se o acesso *on-line* gratuito a artigos avaliados por pares e também a pré-publicações não revisadas, permitindo ao leitor baixá-los, copiá-los, distribuí-los, imprimir-los ou reutilizá-los sem nenhuma

restrição econômica. Nesse sentido, a função do *copyright se resume ao* reconhecimento de autoria do trabalho intelectual e ao direito de citação do autor.

Além disso, a Iniciativa de Acesso Aberto de Budapeste (2002) destaca a adoção de duas estratégias complementares para alcançar o acesso aberto: auto-arquivamento e periódicos de acesso aberto. A primeira estratégia refere-se ao depósito de artigos em repositórios eletrônicos abertos. Por seu turno, a segunda diz respeito à atitude de periódicos eletrônicos concederem acesso aberto permanente aos artigos que publicam, o que significa a extinção de taxas de acesso, ou cobranças de assinaturas.

A Declaração de Bethesda foi redigida durante reunião realizada em abril de 2003, na sede do *Howard Hughes Medical Institute* em Chevy Chase, Maryland. Esse documento teve o objetivo de ampliar as discussões na área da Biomedicina sobre o acesso aberto à literatura, de modo a incluir o ponto de vista de diferentes atores da comunicação científica, tais como agências de fomento à pesquisa, pesquisadores e bibliotecários (BETHESDA STATEMENT ON OPEN ACCESS PUBLISHING, 2003).

Segundo a declaração mencionada, uma publicação de acesso aberto é aquela que cumpre os dois requisitos seguintes (BETHESDA STATEMENT ON OPEN ACCESS PUBLISHING, 2003):

1. o autor e o possessor dos direitos autorais concedem o direito gratuito, irrevogável, mundial e ininterrupto de acesso aos trabalhos científicos, de modo que não haja restrições relacionadas a seu uso, distribuição e exibição pública para qualquer finalidade responsável, desde que ocorra a devida atribuição autoral;
2. logo após a publicação inicial, uma versão completa do trabalho e os materiais suplementares decorrentes da pesquisa são depositados em um repositório *on-line*. Esse, geralmente é mantido por instituição de ensino e pesquisa que garante acesso gratuito e viabiliza o arquivamento de documentos a longo prazo.

Publicada em 2003, a Declaração de Berlim sobre acesso ao conhecimento nas Ciências e Humanidades anuncia a *Internet* como instrumento para a disseminação do conhecimento científico em nível global, possibilitando a adoção de medidas em

favor do acesso aberto por parte de institutos de pesquisa, instituições financiadoras, decisores políticos e unidades de informação, como museus, bibliotecas e arquivos. Toma-se nota, ainda, que a declaração amplia o conceito de acesso aberto ao incluir, além de publicações tradicionais, o acesso a fontes de informação, tais como, resultados de pesquisas originais, dados desestruturados e metadados, e materiais de multimídia (DECLARAÇÃO DE BERLIM, 2003).

Os Princípios *Panton* (2010) empenham-se na defesa e promoção do acesso aberto a dados de pesquisa. Esses, por sua vez, podem ser entendidos como dados gratuitamente disponíveis na *Internet*, passíveis de serem baixados, copiados e transmitidos a quaisquer usuários, sem restrições de proveniência financeira, legal ou técnica. Os Princípios *Panton* são (MURRAY-RUST *et al.*, 2010):

- 1. ao publicar dados, faça uma declaração explícita e robusta de suas pretensões** – ao se publicar dados ou conjuntos de dados, é indispensável a elaboração de uma declaração irrevogável que explicita as pretensões de editores em relação ao uso desses dados;
- 2. utilize uma licença apropriada aos dados** – uso de licenças projetadas para o contexto dos dados de pesquisa;
- 3. se você deseja que seus dados sejam efetivamente usados e incorporados em pesquisas de terceiros, o acesso a esses dados não pode ser realizado mediante cláusulas de licenças restritivas** – o uso de licenças limitantes cerceiam a produção de pesquisas derivadas e dificultam o reaproveitamento eficaz de conjunto de dados;
- 4. as licenças PDDL (*Public Domain Dedication and Licence*) ou CCZero (*Creative Commons Zero Waiver*) são altamente recomendáveis para a promoção do acesso aberto aos dados** – principalmente àqueles dados obtidos por meio de financiamento público.

Reconhecendo o valor estratégico e informacional proporcionados pelo amplo acesso a dados de pesquisa, o Manifesto de acesso aberto a dados da pesquisa brasileira para ciência cidadã (publicado em 2016, pelo Instituto Brasileiro de Informação em Ciência e Tecnologia – IBICT) demonstra a importância de dados

serem compartilhados, revisados, redistribuídos e reutilizados em novos contextos da produção científica.

O manifesto destina-se a vários atores que integram o sistema de comunicação científica, incluindo: institutos de pesquisa, universidades, sociedades científicas, agências de fomento à pesquisa, editores de periódicos científicos; programas de pós-graduação e cursos de graduação nas áreas de informação, gestores de dados de pesquisa e pesquisadores (IBICT, 2016).

No que se refere ao acesso aberto a dados de pesquisa, o manifesto recomenda às universidades e institutos de pesquisa brasileiros a observação das seguintes diretrizes (IBICT, 2016):

- elaborar políticas institucionais de caráter mandatório para garantir que dados de pesquisa sejam devidamente gerenciados e disponibilizados por meio de acesso aberto;
- incentivar a realização de pesquisas colaborativas e interdisciplinares;
- integrar dados de pesquisa à memória acadêmica;
- apoiar ações para o desenvolvimento de repositórios de dados de pesquisa;
- utilizar *hardware* e *software* abertos;
- recomendar a elaboração de plano de gestão de dados de pesquisa a pesquisadores e profissionais de informação;
- favorecer o acesso aos dados de pesquisa à sociedade como um todo.

Bibliotecários latino-americanos reunidos na *Conferencia Latinoamericana em Ciencia Abierta* (OpenCon LatAm, 2019), realizada em Bogotá, reconheceram a ciência aberta como alternativa viável para a adoção de um modelo sustentável que assegure a gestão e o compartilhamento de dados de pesquisa à toda a sociedade, respeitando sua diversidade. Como atores-chave para incentivar uma mudança cultural em direção à Ciência aberta, esses profissionais elaboraram o intitulado: Manifesto bibliotecário pela Ciência Aberta na América Latina (2019), comprometendo-se em desenvolver e apoiar as seguintes ações (MANIFIESTO BIBLIOTECARIO POR LA CIENCIA ABIERTA EN AMÉRICA LATINA, 2019):

1. criar estratégias para a construção de uma rede composta por diversos atores a favor de práticas de abertura da pesquisa latino-americana;

2. incentivar o trabalho colaborativo entre bibliotecários, de modo a disseminar boas práticas na gestão e comunicação do conhecimento científico e cultural;
3. estimular a elaboração de políticas públicas institucionais de Ciência aberta;
4. criar programas de capacitação para promover a Ciência aberta;
5. fortalecer o uso de infraestruturas de código aberto e interoperáveis;
6. conduzir a gestão institucional de dados de pesquisa, por meio da formulação de políticas de dados de pesquisa, criação de planos de dados, entre outras ações, a fim de armazenar, descrever, organizar e preservar esses dados;
7. criar serviços para apoiar a pesquisa científica;
8. apoiar ações relacionadas à ciência cidadã;
9. sensibilizar comunidades de usuários para o entendimento da Ciência aberta;
10. propor novos métodos para medir e avaliar a produção científica, de forma que práticas abertas sejam consideradas nesses processos.

Segundo Mueller (2006, p. 27) “o movimento para acesso livre ao conhecimento científico pode ser considerado como o fato mais interessante e talvez importante de nossa época no que se refere à comunicação científica”. Considerando a relevância desse tópico, a sessão seguinte dedica-se à apresentação de algumas definições e conceitos relacionados ao acesso aberto, de modo a favorecer o entendimento do contexto no qual esse movimento se situa e a difundir algumas de suas vantagens no âmbito da comunicação científica.

2.2.1.1 O que é o acesso aberto?

“Uma forma revolucionária de acesso” foi a expressão utilizada por Suber (2015, p. 58, *tradução nossa*) para caracterizar o movimento denominado: Acesso aberto. Sob esse enfoque, a literatura em acesso aberto é aquela que está disponível em meio digital e *on-line*, gratuitamente, e livre de restrições – econômicas ou relacionadas ao seu reuso.

Restrições econômicas demonstram que o preço é uma barreira considerável a pesquisadores, professores e estudantes que necessitam acessar diferentes obras para o desenvolvimento de projetos de pesquisa e às bibliotecas que empenham

recursos para a aquisição de centenas de títulos para a formação e desenvolvimento de suas coleções. Já as restrições relacionadas ao uso, referem-se a limitações no âmbito dos direitos autorais. Nesse sentido, há textos os quais só é possível realizar a sua leitura, inviabilizando intervenções, tais como, a tradução para outros idiomas, a modificação de formatos compatíveis às novas tecnologias, a mineração e o processamento de dados (SUBER, 2015).

Segundo Melero (2005), o acesso aberto a publicações científicas está relacionado a iniciativas que viabilizam acesso livre e sem barreiras à literatura produzida pela comunidade científica. Anular tais barreiras favorece a visibilidade e a disseminação de pesquisas, oferecendo um retorno à sociedade, principalmente em relação a estudos realizados a partir de financiamento público. Portanto, o acesso aberto contribui para “melhorar o sistema tradicional de comunicação científica e facilitar o livre acesso às publicações científicas por meio da Internet” (MELERO, 2005, p. 258, *tradução nossa*).

Semelhantemente, Alonso Arévalo, Subirats Coll e Martínez Conde (2008, p. 20, *tradução nossa*) afirmam que “o movimento de acesso aberto objetiva melhorar a comunicação científica e eliminar todos os tipos de barreiras que impedem o acesso à informação, maximizando o acesso à mesma por meio do auto-arquivamento”.

De acordo com Leite (2009), Stevan Harnad – um dos mais influentes pesquisadores precursores e incentivadores do movimento do Acesso aberto – defende a implantação de duas vias estratégicas para conduzir o acesso aberto à literatura científica: a via dourada e a via verde. Por um lado, entende-se pela via dourada a produção de periódicos de acesso aberto disponíveis na rede. A publicação de artigos por meio dessa via potencializa o diálogo entre pesquisadores, derruba barreiras econômicas enfrentadas por unidades de informação relacionadas à aquisição de novos títulos e acelera o fluxo da comunicação científica, tornando mais célere a disseminação de novidades que podem potencializar avanços científicos. Para tanto, várias ferramentas foram desenvolvidas com o propósito de viabilizar a produção de periódicos de acesso aberto. Essas, por sua vez, são construídas em *softwares* livres e agilizam trâmites inerentes ao processo editorial.

Por outro lado, a via verde refere-se à criação de repositórios institucionais (RIs) que estruturam, armazenam e disseminam a informação científica produzida por instituições de pesquisa. Os RIs possibilitam a disseminação de diferentes tipos de documentos, incluindo artigos de periódicos, teses e dissertações avaliadas pelos pares (LEITE, 2009). Acrescente-se, oportunamente, as considerações de Suber (2015) nas quais repositórios de acesso aberto beneficiam autores e centros de pesquisas por meio do fornecimento de URLs (*Uniform Resource Locator*, ou Localizador Uniforme de Recursos, em português) permanentes a documentos e da preservação de registros científicos, a longo prazo.

Assim também, Mueller (2006) explica que os RIs cumprem o propósito de reunir documentos produzidos em uma instituição. A título de exemplo, repositórios de universidades reúnem a produção científica gerada nas próprias instituições, disponibilizando-a por meio de coleções de documentos digitais. Geralmente, são incumbidos de resguardar e preservar a memória institucional, ampliando a visibilidade e o acesso a pesquisas produzidas nas universidades. Ademais, os RIs proporcionam acesso aberto a vários tipos de documentos, como trabalhos de docentes apresentados em congressos e seminários, relatórios de pesquisa, teses e dissertações de discentes.

De acordo com Alonso Arévalo, Subirats Coll e Martínez Conde (2008), a publicação de pesquisas em acesso aberto oferece inúmeras vantagens no âmbito da comunicação científica, as quais podem-se destacar as seguintes:

- a pesquisa torna-se acessível a todos os cientistas e seus resultados são rapidamente disseminados;
- artigos de periódicos científicos podem ser indexados por várias bases de dados e serviços de alerta de informações, aumentando a possibilidade de serem lidos e citados;
- é possível medir o impacto de publicações e aplicar diferentes métricas por meio da correlação entre o número de *downloads* de documentos e a quantidade de citações;
- possibilidade de acesso ao texto completo de documentos;

- o acesso aberto é uma alternativa viável que faz uso das tecnologias disponíveis para ampliar a capacidade de distribuição da informação científica. Dessa maneira, reduz custos, com o propósito de ampliar e facilitar o acesso a pesquisas científicas.

Convictos de que o acesso aberto impulsiona a atualização do conhecimento científico de forma democrática, participativa e inclusiva, e influenciados pelas contribuições de Suber em relação a esse movimento, Banerjee, Babini e Aguado (2015) expõem vários argumentos que demonstram a importância e vantagens da adoção do acesso aberto para a comunidade científica e para a sociedade em geral, principalmente no contexto da América Latina e Caribe.

A primeira vantagem a ser destacada é a de que o acesso aberto, por meio de recursos da *Internet*, viabiliza a livre circulação de publicações e metadados, mediante disponibilização de conteúdos em plataformas conectadas em rede. Considerando o contexto tecnológico no qual é possível armazenar e compartilhar eficientemente conteúdos científicos, Banerjee, Babini e Aguado (2015) ainda explicam que isso pode estimular a disseminação de publicações científicas de autorias latino-americanas e favorecer a transmissão de conhecimentos de forma mais democrática e inclusiva, como é possível observar nos seguintes trechos:

A internet se tornou um espaço de confluência universal entre atores que trocam informações por meio de diversas plataformas e dispositivos tecnológicos conectados online, o aa [acesso aberto] tem encontrado nesta um espaço de implementação extremamente eficiente para armazenar, compartilhar e retroalimentar diversos conteúdos de caráter acadêmico, facilitando assim a formação de vasos comunicantes que contribuem para a ampliação do raio de disseminação social da ciência escrita. [...] O aa [acesso aberto] permite a livre mobilização das publicações e de seus metadados em uma ampla gama de buscadores e de plataformas interconectadas na rede, o que favorece a criação de canais de disseminação dos conteúdos científicos produzidos por uma maior diversidade de atores latino-americanos, que impulsionam a inovação, enriquecem o debate acadêmico e melhoram tanto a transferência como a aplicação do conhecimento no meio social de acordo com configurações mais democráticas e inclusivas (BANERJEE; BABINI; AGUADO, 2015, p. 16, *tradução nossa*).

Após a narrativa de estudos acerca do acesso aberto, a seção seguinte dedicar-se-á à apreciação de estudos na literatura científica e técnica que versam sobre os dados de pesquisa, uma vez que o presente trabalho está intimamente

ligado a esse tópico. Sendo assim, serão abordados assuntos acerca da tipologia de dados, dos dados abertos de pesquisa e de licenças para conjuntos de dados abertos.

2.3 Dados de pesquisa

De acordo com Borgman (2015), apesar de o termo “dados” ter sido discutido ao longo de pelo menos cinco séculos, ainda não há um consenso a seu respeito, uma vez que a sua definição varia significativamente conforme o usuário e o contexto os quais se referem. Para fim de discussão, em sua obra intitulada *Big Data, little data, no data: scholarship in the networked world*, Borgman (2015, p. 29, tradução nossa, grifo da autora) afirma que “os dados referem-se a entidades usadas como evidência de fenômenos para fins de pesquisa”. Logo, essas entidades tornam-se dados somente quando alguém as consideram evidências de um fenômeno. A título de ilustração, a autora demonstra que fotografias de um álbum de família ou um anuário do ensino médio podem se tornar dados de pesquisa para um pesquisador que os utilize como evidências de traços históricos e culturais de estilos de cabelos e roupas de determinada época. Por outro lado, um pesquisador também poderia utilizá-los em estudos que explorem os agrupamentos familiares ou identidades sociais. Outrossim, antigos registros meteorológicos para fins de navegação, hoje, podem ser utilizados como dados de pesquisa relacionados ao estudo de alterações climáticas.

Acrescente-se à definição de dados de pesquisa apresentada por Borgman (2015), a noção de validação de resultados de pesquisas demonstrada no relatório: *Principles and guidelines for access to research data from public funding*, elaborado pela *Organisation For Economic Co-operation And Development* (OECD), e em estudos de Sayão e Sales (2014). À vista disso, os dados podem ser entendidos como registros factuais utilizados nas pesquisas, comumente aceitos na comunidade científica para atribuir veracidade e autenticidade a resultados obtidos.

É interessante notar que para Borgman (2015), os dados de pesquisa não possuem valor ou significado quando compreendidos em contexto isolado. Ao contrário, o significado desses dados encontra-se dentro de uma infraestrutura de conhecimento constituída por uma ecologia de pessoas, instituições, tecnologias, entre

outros elementos. Considerando o exposto, a autora explica a diversidade de dados de pesquisa em razão de diferenças entre áreas do conhecimento. Nessa perspectiva, domínios do conhecimento diferem-se principalmente quanto aos tipos de dados produzidos ou obtidos, às práticas de compartilhamento e reuso de dados, à definição de critérios para atribuição de autoria para dados, ao modo pelo qual dados são interpretados, entre outros aspectos. Sobre a diversidade de dados, Borgman (2015, p. 79) afirma:

A diversidade de dados deve-se não apenas à amplitude das abordagens de pesquisa nos domínios da investigação, mas também às muitas maneiras pelas quais observações individuais, objetos, registros, textos, espécimes, pessoas, animais ou outras entidades podem ser representados como dados. Os campos de estudo diferem por esses e muitos outros fatores, tais como os objetivos de seus projetos de pesquisa, as maneiras pelas quais dados são coletados e analisados e as escolhas de novas fontes ou recursos existentes para dados (...) a mesma entidade pode ser representada de várias maneiras, por muitos meios e para muitos propósitos, por longos períodos de tempo. No processo de representação, essas entidades podem adquirir muitos nomes e esses nomes podem ser organizados de várias maneiras (BORGMAN, 2015, p. 79, *tradução nossa*).

Considerando que dados de pesquisa são entidades intrínsecas à produção do conhecimento científico no contexto de universidades, serão destacadas explicações sobre esse termo elaboradas por algumas dessas instituições. A título de exemplo, serão detalhadas as definições de seis universidades, a saber: Universidade de Leeds; *University College London (UCL)*; Universidade de Nottingham; Universidade de Yale; Universidade de São Paulo; Universidade de Melbourne.

Para a Universidade de Leeds⁸, dados são quaisquer informações coletadas para fins de comprovação dos resultados originais de uma pesquisa. Disponíveis em diferentes formatos, os dados de pesquisa podem incluir documentos, planilhas, cadernos de laboratório, diários de pesquisa, questionários, transcrições, áudios, vídeos, fotografias, respostas de testes, *slides*, espécimes, artefatos, modelos, arquivos de dados, algoritmos, fluxos de trabalho, protocolos, metodologias, entre outros.

8 Disponível em:
https://library.leeds.ac.uk/info/14062/research_data_management/61/research_data_management_explained

Ainda segundo a Universidade de Leeds, dados de pesquisa são gerados a partir de diferentes fontes, com finalidades distintas. Tendo em vista esses aspectos, podem ser classificados em:

- **dados observacionais** – geralmente insubstituíveis, são dados capturados em tempo real, como dados coletados por meio de sensores, imagens neurais e dados coletados por meio de *survey*;
- **dados experimentais** – dados capturados no ambiente de laboratório e frequentemente replicáveis, tais como o sequenciamento genético e cromatogramas;
- **dados de simulação** – dados originados a partir de modelos de testes, como modelos climáticos e modelos econômicos;
- **dados derivados ou compilados** – originários de dados já existentes, são dados que foram transformados ou adaptados em razão de determinado contexto de pesquisa. Dados decorrentes de mineração e modelos 3D ilustram essa categoria de dados;
- **dados de referência ou canônicos** – são coleções de dados que muito provavelmente foram revisados por pares, publicados e passaram por tratamento técnico em relação a seu arquivamento e disponibilização. Como exemplo, pode-se destacar os bancos de dados de sequenciamento genético e estruturas químicas.

De maneira semelhante, a política de dados de pesquisa elaborada pelo *UCL Research Information and IT Services Group (RIISG)* da *University College London* define os dados de pesquisa como fatos sobre os quais um argumento ou teoria é elaborado ou testado.

Por seu turno, a Universidade de Nottingham⁹ define o termo “dados de pesquisa” como tudo o que é coletado com a finalidade de produzir resultados originais de pesquisa. Isso inclui: gravações de vídeo, correspondências, fitas de áudios, procedimentos operacionais, entre outros. De modo semelhante, esse termo

9 Disponível em: <https://www.nottingham.ac.uk/fabs/rgs/research-data-management/what-is-research-data.aspx>

é entendido pela Universidade de Yale¹⁰ como a informação coletada com o propósito de analisar a produção de pesquisa original. Logo, dados de pesquisa podem ser utilizados na construção de argumentos científicos. Além disso, podem ser derivados de observações, como dados pluviométricos, decorrentes de levantamento, coletados por meio de simuladores, obtidos no ambiente de laboratórios ou por meio de mineração de textos, entre outros.

De acordo com informações retiradas do *site* da Universidade de São Paulo¹¹, os dados são registros que embasam resultados de pesquisas. Os dados de pesquisa diferem-se de acordo com as áreas do conhecimento e, em razão disso, podem incluir documentos textuais, planilhas, conjuntos de dados estatísticos, cadernos de laboratório, diários de campo, cadernos de campo, diários, questionários, sequências genéticas, artefatos e outros materiais. Apesar de dados exercerem funções fundamentais na produção do conhecimento científico, ainda há muita dificuldade em estruturá-los, descrevê-los, disponibilizá-los e preservá-los a longo prazo.

Segundo a política de gestão de dados de pesquisa elaborada pela biblioteca da Universidade de Melbourne (2013), os dados de pesquisa referem-se a diversos contextos, estando disponíveis em suportes e formatos variados. Em síntese, esses dados podem ser definidos como fatos nos quais se baseiam argumentos ou teorias. Acrescentam-se, oportunamente, as considerações feitas a respeito de dados de pesquisa no âmbito das artes. Sob essa perspectiva, os dados incluem artefatos que representam a expressão humana por meio das artes plásticas e cênicas, da música, e outras formas artísticas.

2.3.1 Tipologia de dados

De acordo com Silva (2019), uma primeira distinção tipológica em relação aos dados pode ser realizada levando em consideração a sua natureza quantitativa ou qualitativa. Em geral, dados quantitativos são expressos por meio de algarismos numéricos. Por outro lado, os dados qualitativos referem-se à descrição de fenômenos,

10 Disponível em: <https://guides.library.yale.edu/datamanagement>

11 Disponível em: <http://www.sibi.usp.br/apoio-pesquisador/dados-pesquisa/>

atribuindo a esses, qualidades em abordagem de categorização, e não quantificação. Sob esse ponto de vista, pesquisadores no campo da Matemática seriam mais propensos a utilizar dados quantitativos, enquanto pesquisadores na área das humanidades estariam mais inclinados a fazer uso de dados qualitativos. Todavia, o autor destaca que tal distinção não é tão óbvia o quanto parece ser, na medida em que diferentes áreas do conhecimento utilizam ambos os tipos de dados na condução de suas pesquisas.

Com base na tipologia de dados apresentada pela *National Science Foundation* (NSF, 2007), cujos dados são categorizados por critérios relativos aos procedimentos os quais são coletados, ao caráter primário, secundário ou terciário que possuem, ao nível de estruturação e aos formatos em que são disponibilizados, Silva (2019) comenta tais categorias com o propósito de apresentar um panorama sobre a vasta tipologia de dados de pesquisa, como pode ser visto nos seguintes tópicos:

Quanto aos procedimentos de coleta, os dados podem ser classificados em:

- **dados observacionais** – dados capturados em tempo real, geralmente não podendo ser replicados ou substituídos, na medida em que possuem caráter único. A título de ilustração, pode-se mencionar os registros climáticos e de fauna e flora coletados por sensores, satélites ou por anotações, além de registros de observações sobre o comportamento de determinado grupo de pessoas;

Considerando o contexto dos dados observacionais e a pertinência do assunto, é interessante fazer menção aos dados geográficos. Segundo a empresa de *software Heavy.ai*¹², sediada nos Estados Unidos, geodados, também chamados de dados geográficos ou dados geoespaciais, referem-se a dados e informações sobre localizações geográficas, ou seja, dizem respeito a um local em relação à Terra. Englobam arquivos vetoriais, bancos de dados geográficos, dados multitemporais, entre outras fontes. Além disso, podem ser obtidos por diversos meios, incluindo dispositivos telemáticos, *Global Positioning System*, imagens geoespaciais coletadas via satélite e Internet das coisas.

- **dados computacionais** – dados resultantes de modelos de ordenações, simulações ou fluxos de processos passíveis de serem reproduzidos, desde que preservados em integridade. As áreas da Bioinformática e Genômicas foram precursoras na utilização de dados computacionais. Esses dados são comumente relacionados ao contexto das Ciências Físicas e Biológicas, apesar de uso na condução de pesquisas nos campos de Ciências Sociais e Humanidades. A título de exemplo, pode-se mencionar a utilização de dados computacionais por oceanógrafos no estudo de marés e por economistas em análises sobre interações de mercado;
- **dados experimentais** – dados resultantes de procedimentos experimentais, ou seja, realizados em ambiente controlado com o propósito de validar ou refutar hipóteses a respeito de determinado fenômeno.

Quanto ao caráter primário, secundário ou terciário dos dados de pesquisa, Silva (2019) explica que:

- **dados primários** – são derivados de trabalhos originais, sendo coletados para atender aos objetivos previamente estabelecidos em determinada pesquisa;
- **dados secundários** – são dados disponíveis publicamente em diversas fontes de informação, tais como livros, periódicos científicos e censos disponibilizados em repositórios e bases de dados, e que podem ser utilizados em outros estudos com objetivos distintos;
- **dados terciários** – são dados derivados de outras pesquisas que incluem procedimentos como recontagens, categorizações e resultados estatísticos. Com frequência, são dados utilizados para assegurar a confidencialidade de dados primários ou secundários.

À vista disso, é oportuno destacar os comentários de Silva (2019) nos quais foi apontado que, geralmente, a obtenção de dados primários requer um custo e tempo maior que a obtenção de dados secundários. Entretanto, a combinação de dados primários a dados secundários e terciários pode gerar dados ainda mais valiosos ao contexto científico. Isso porque esses dados possibilitam a reprodução de resultados de pesquisa e a criação de conjuntos de dados combinados que refletem associações e padrões os quais não são facilmente percebidos.

Quanto ao nível de estruturação que apresentam, os dados podem ser classificados em (SILVA, 2019):

- **dados estruturados** – estão armazenados de modo estruturalmente identificável, podendo ser transferidos a outros sistemas sem maiores dificuldades. Geralmente, esses dados estão dispostos em tabelas ou bases de dados relacionais e são organizados mediante esquema de estruturação;
- **dados semiestruturados** – abrangem imagens e documentos textuais, os quais não há um modelo de organização ou esquema estrutural predeterminado. Dados contidos em páginas *web* exemplificam esse tipo de dado;
- **dados não estruturados** – não possuem uma estrutura identificável e, por essa razão, são de difícil combinação e análise em termos de recuperação da informação em ambiente *web*. Em geral, são de natureza qualitativa, sendo convertidos em dados estruturados por meio de procedimentos de classificação e categorização. Nesse contexto, muitos estudos são conduzidos sobre dados não estruturados contidos em mensagens de redes sociais e *blogs*.

Com relação aos formatos de arquivos de dados, Silva (2019) explica que o formato e o *software* pelos quais dados de pesquisa são criados resultam dos modos como pesquisadores produzem e analisam esses dados, sendo diretamente influenciados por tradições das disciplinas em que atuam. Considerando a produção de dados em meio digital, e que não há diretrizes específicas para a escolha de formatos de arquivo, sugere-se que pesquisadores optem, preferencialmente, pela adoção de formatos abertos, padronizados e documentados. Por outro lado, recomenda-se, sempre que possível, evitar formatos de arquivos desconhecidos ou proprietários. Nesse sentido, o Quadro 2 esquematiza exemplos dos principais formatos de arquivos aceitos para intercâmbio, reutilização e preservação, e de outros formatos aceitáveis para a preservação de dados, de acordo com os tipos de dados.

Quadro 2 – Formatos de arquivos

TIPO DE DADOS	PRINCIPAIS FORMATOS ACEITOS PARA INTERCÂMBIO, REUTILIZAÇÃO E PRESERVAÇÃO	OUTROS FORMATOS ACEITÁVEIS PARA A PRESERVAÇÃO DE DADOS
Dados tabulares quantitativos com metadados ampliados Um conjunto de dados com etiquetas variáveis, etiquetas de códigos e os valores que faltam definir, assim a matriz de dados.	Formato portátil SPSS (.por). Texto e comando arquivo delimitado (<i>setup</i>) (SPSS, Stata, SAS, etc.) que contém a informação de metadados. Um texto estruturado que contém informação de metadados margem de arquivo, por exemplo, Arquivo XML DDI.	Formatos proprietários de pacotes estatísticos, exemplo SPSS (.sav), Stata (.dta), MS Access (.mdb/accdb).
Dados tabulares quantitativos com metadados mínimos Uma matriz de dados com ou sem descrição de coluna ou nomes de variáveis, mas não outros metadados ou etiquetado.	Valores separados por vírgulas (CSV) (.csv). Arquivo delimitado por tabuladores (TAB) incluindo texto delimitado de carácter determinado conjunto com instruções de definição de dados SQL.	Texto delimitado de caracteres de dados (únicos personagens que não estão presentes nos dados); se deve utilizar como delimitadores (.txt). Formatos amplamente utilizados, por exemplo, MS Excel (.xls/xlsx). MS Access (.mdb/accdb), dBase (.dbf) e folhas de cálculo de Open Document (ODS). Texto com caracteres delimitados (.txt).
Dados geoespaciais Dados vetoriais e raster.	ESRI Shapefile (essencial: shp, .shx, .dbf; opcional: prj, .sbx, .sbn). TIFF georreferenciada (tif, tfw). Dados CAD (dwg). Dados de atributos GIS tabular. Keehole Markup Language (KML) (kml).	Formato ESRI Geodatabase (mdb). MapInfo Interchange (mif) para dados de vetor. Adobe Illustrator (ai), dados CAD (dxf o .svg). Formatos binários de pacotes GIS e CAD.
Dados qualitativos Textual.	eXtensible Markup Linguagem (XML), utilizado de acordo com um tipo de documento apropriado (DD) o esquema (.xml). Formato de texto enriquecido (.rtf). Os dados de texto sem formato, ASCII (txt), UTF-8 (Unicode).	HyperText Markup Language (HTML) (.html). Formatos amplamente utilizados, por exemplo, MS Word (doc/docx). Alguns formatos proprietários/específicos de software, por exemplo, NUD * IST, NVivo e ATLAS. Ti LaTeX (.txt).
Dados de imagem digital	TIFF versão 6 sem comprimir (.tif).	JPEG (.jpeg, .jpg), mas somente se cria neste formato. TIFF (outras versões) (tif, .tiff). Adobe Portable Document Format (PDF/A, PDF) (pdf). Norma RAW aplicável ao formato de imagem (.raw). Arquivos de Photoshop (.psd).
Dados de áudio digital	Free Lossless Audio Codec (FLAC) (.flac). Formato de forma de onda de áudio (WAV) (.wav), Waveform Audio Format.	MPEG-1 Áudio Laer 3 (.mp3) mas somente se cria neste formato. Audio Inter Change File Format (AIFF) (.aif).
Dados de vídeo digital	MPEG-4 (.mp4).	JPEG 2000 (.mj2).
Documentação e plano de gestão dos dados	Formato de texto enriquecido (.rtf), Rich Text Format. HTML (.htm, .html). Open Document Text (.odt).	Texto plano (txt). Alguns formatos proprietários amplamente utilizados, por exemplo, MS Word (doc/docx) o MS Excel (.xls/xlsx). XML marcador de texto (.xml) de acordo com um DD ou esquema adequado, por exemplo, XHML 1.0. PDF (.pdf).

Fonte: Silva (2019, p. 39-40), adaptado das recomendações da *UK Data Archive*

2.3.2 Dados abertos

Oportunamente, acrescente-se à seção relativa aos dados de pesquisa, a subseção que aborda os dados abertos, os quais são definidos pela *York University*¹³

13 Disponível em: <https://www.library.yorku.ca/web/open/overview/open-data/>

como aqueles dados que estão disponíveis na *Internet* gratuitamente, possibilitando a qualquer pessoa o seu *download*, alterações e distribuição, sem quaisquer limitações financeiras ou legais. Logo, os dados abertos caracterizam-se por ser:

- disponíveis – possibilidade de dados serem baixados por meio da *Internet*;
- acessíveis – disponibilizados em formatos passíveis de modificação;
- reutilizáveis – desde que explicitamente disposto nos termos de uso;
- redistribuíveis – podendo ser combinados a dados derivados de outras pesquisas;
- sem restrições – qualquer pessoa pode utilizar, alterar e compartilhar os dados, irrestritamente, incluindo fins comerciais, não-comerciais e educacionais.

Segundo a *York University*, os dados abertos são importantes por fornecerem diversos benefícios à comunidade de pesquisadores, dentre os quais pode-se destacar o aumento da reprodutibilidade da pesquisa e das oportunidades de colaboração entre cientistas, além de evitar a duplicação de estudos. Para tanto, é imprescindível que tais dados sejam interoperáveis, na medida em que diversos pesquisadores e instituições poderão compartilhá-los, e, conseqüentemente, trabalhar em colaboração.

Conforme o entendimento de Silva (2019, p. 33), os dados abertos de pesquisa são aqueles que “podem ser reutilizados e distribuídos por qualquer pessoa, sem barreiras técnicas, legais, sob os requerimentos de reconhecer a autoria e compartilhar o novo produto nas mesmas condições”. Para tanto, o autor explica que o compartilhamento público de dados de pesquisa deve ocorrer mediante a escolha de pesquisadores em relação à licença de direitos de reutilização, que estabelece o que é permitido e o que é proibido nesse contexto.

Dentre os vários tipos de licenças que dizem respeito ao contexto eletrônico e de padrões para conjuntos de dados abertos, as licenças *Open Data Commons* e *Creative Commons* são as mais utilizadas. A primeira refere-se à iniciativa promovida pela *Open Knowledge Foundation* (OKF) e abrange vários instrumentos jurídicos que auxiliam a criação e uso de dados abertos. Por seu turno, a segunda cumpre o objetivo de estimular pesquisadores a definir as condições e termos os quais suas obras serão utilizadas (SILVA, 2019).

Conforme definido em manual sobre dados de pesquisa elaborado pela OKF¹⁴, dados abertos são dados que podem ser livremente utilizados, reutilizados e disseminados por quaisquer indivíduos, exigindo-lhes, ao máximo, a atribuição de fonte e o compartilhamento de dados pelas mesmas prescrições. À luz desse contexto, quatro aspectos são destacados no manual, os quais são resumidamente apresentados nos seguintes tópicos:

- **disponibilidade de acesso** – dados devem ser preferencialmente disponibilizados na *Internet* de modo conveniente e modificável;
- **reutilização e redistribuição** – dados devem ser disponibilizados sob condições que permitam sua reutilização e redistribuição, incluindo a integração com outros conjuntos de dados;
- **participação universal** – a disponibilização de dados deve ocorrer sem restrição ou discriminação a determinada área de atuação, pessoa ou grupo;
- **interoperabilidade** – possibilita a integração entre diferentes conjuntos de dados, permitindo que diversos sistemas ou instituições trabalhem cooperativamente.

Segundo a OKF¹⁵, os dados abertos podem ser classificados em relação ao seu potencial uso e aplicação, como sendo:

- **culturais** – referem-se a trabalhos culturais e artefatos geralmente coletados e armazenados em unidades de informação, tais como galerias de artes, bibliotecas, arquivos e museus;
- **científicos** – dados produzidos para fins da pesquisa científica;
- **financeiros** – incluem dados relativos a contas governamentais, tais como despesas e receitas, além de informações referentes ao mercado financeiro, como ações e títulos;
- **estatísticos** – referem-se a dados estatísticos, tais como àqueles relacionados a censos e indicadores socioeconômicos;
- **climáticos** – incluem diferentes tipos de registros relativos à compreensão e previsão de condições climáticas;

14 Disponível em: http://opendatahandbook.org/guide/pt_BR/what-is-open-data/

15 Disponível em: <https://okfn.org/opendata/>

- **ambientais** – dados relacionados com o ambiente natural, tais como o nível de poluentes em determinado ambiente e da qualidade da água em rios e mares.

Acrescente-se a esse contexto, os dados abertos governamentais, mencionados em manual de dados abertos da OKF¹⁶. Esse tipo de dado refere-se à produção governamental, que, por sua vez, abrange uma enorme quantidade de dados ainda pouco explorada. Todavia, dados abertos governamentais contribuem para a transparência e controle sobre a atuação governamental, melhorias na eficiência e efetividade de serviços públicos, verificação do impacto de políticas públicas e produção de novos conhecimentos mediante a integração de várias fontes de dados.

Considerando aspectos relevantes à discussão sobre dados, Molloy (2011) destaca que o processo de abertura dos mesmos é fundamental ao desenvolvimento científico, em razão de atribuir maior transparência à pesquisa e possibilitar a reprodutibilidade de resultados obtidos. Nesse sentido, a ciência avança a partir de dados que são produzidos ou obtidos em diferentes áreas do conhecimento. Todavia, de acordo com a autora, diversos são os desafios que permeiam a livre disponibilização e acesso a dados de pesquisa. A título de exemplo, Molloy (2011) ressalta as restrições de uso estabelecidas por editores e fornecedores de dados, as dificuldades relativas ao reuso de dados ocasionadas por inconsistências na apresentação e nos formatos de disponibilização desses, e a falta de incentivo aos pesquisadores para o compartilhamento aberto dos dados que produzem.

Tendo em vista a relevância dos dados de pesquisa no contexto científico destacada por Molloy (2011), toma forma o questionamento sobre quais caminhos devem ser percorridos por pesquisadores em direção à abertura dos dados que produzem ou obtêm. Nesse sentido, a Agência para a Modernização Administrativa (AMA) – instituto público responsável pela modernização de aspectos administrativos em Portugal – elaborou em 2016, o “Guia de dados abertos”. Essa publicação aponta cinco fases fundamentais para que dados sejam efetivamente abertos por instituições. Essas fases incluem o devido planejamento para o processo de abertura de dados,

16 Disponível em: <https://opendatahandbook.org/guide/en/why-open-data/>

a definição dos dados que devem ser abertos, a disponibilização efetiva de dados conforme princípios norteadores, a estratégia de divulgação dos dados abertos e a manutenção de dados. A Figura 4 apresenta essas fases, as quais serão abordadas nos tópicos seguintes:

Figura 4 – Cinco fases para a abertura de dados



Fonte: Agência para a Modernização Administrativa (2016, p. 18)

- **planejamento** – a disponibilização de dados abertos requer uma reflexão estratégica sobre diversos aspectos, os quais referem-se à identificação de formatos abertos e reutilizáveis, às condições internas para atualização de dados, quando necessário, e à responsabilização por dados disponibilizados;
- **definição dos dados que devem ser abertos** – inclui a seleção de dados a serem disponibilizados. Geralmente, instituições analisam o perfil e preferências de usuários, o interesse e utilidade social de dados e seu potencial econômico. Nessa direção, é fundamental o envolvimento da comunidade como potencial usuária de dados em várias perspectivas, tais como a de cidadãos, Organizações Não Governamentais (ONGs), pesquisadores, entre outros. Recomenda-se, então, identificar as necessidades da comunidade em relação à reutilização de dados, a partir de sugestões para a composição de bases de *datasets*, as quais podem ser obtidas por meio de consultas públicas na *Internet* e pelo correio eletrônico;
- **disponibilização efetiva de dados** – garantir suporte à publicação de dados de pesquisa é fundamental ao processo de abertura desses. A começar, isso pode ser realizado por meio de soluções menos complexas, como a disponibilização de dados em *sites* institucionais, ou pode avançar no sentido de soluções mais trabalhosas, como a disponibilização desses dados em plataformas de catálogo de dados abertos. Tais plataformas, por sua vez, consistem em ferramentas *web* que viabilizam o acesso aberto de *datasets* a usuários, possibilitando

pesquisas de dados disponíveis, *downloads* de dados abertos; consultas a metadados, entre outras funcionalidades;

- **estratégia de divulgação dos dados abertos** – a divulgação nesse sentido é importante, principalmente, para alertar potenciais usuários a respeito da disponibilidade de dados em formatos abertos. Para tanto, instituições podem lançar mão de divulgação em *web sites* e redes sociais, a partir da criação de grupos de discussão, ou utilizando grupos já existentes sobre a temática dos dados de pesquisa;
- **manutenção de dados** – o trabalho em relação à abertura de dados não se conclui na disponibilização dos mesmos e em sua posterior divulgação. Ao contrário, esse é um compromisso institucional ao longo do tempo que perpassa as práticas relativas à disponibilidade, atualização, continuidade e qualidade dos dados. Sendo assim, instituições devem assegurar que dados abertos mantenham-se disponíveis continuamente, sejam atualizados, permaneçam fiáveis e duradouros (principalmente dados que apresentam espectro temporal), sejam disponibilizados em períodos com regularidade previsível e que apresentem qualidade em relação a diversos aspectos, permitindo fácil leitura, manipulação e reuso.

Logo, observa-se que a compreensão dos estágios que viabilizam a abertura de dados é fundamental para garantir o acesso e a preservação desses a longo prazo. Entretanto, a questão basilar a que se faz menção neste momento, refere-se ao entendimento de licenças utilizadas para legalizar a abertura de dados, principalmente, em termos de sua reutilização. À vista disso, a seção seguinte dedicar-se-á à exposição de alguns desses instrumentos legais.

2.3.3 Licenças para conjuntos de dados abertos

De acordo com Sayão e Sales (2015), o *copyright* refere-se ao direito de propriedade intelectual atribuído ao criador. Seu registro proíbe a realização de cópias e publicação de uma obra original não autorizada pelo detentor dos direitos autorais. Em geral, produtos de pesquisa, como relatórios e planilhas, são amparados por leis

de *copyright* por se enquadrarem como obras literárias. Entretanto, convém destacar que *copyright* também se aplica a contexto mais abrangente relacionado a dados de pesquisa, sendo um aspecto importante para nortear legalmente a prática de produção, compartilhamento e reutilização de dados.

Nessa direção, Sayão e Sales (2015) ainda explicam que os detentores do *copyright* mantêm os direitos autorais sobre seus dados, ainda que esses sejam compartilhados ou arquivados. De mesmo modo, para que um repositório archive dados, é necessário que os detentores de seus direitos autorais concedam autorização explícita permitindo o compartilhamento de tais recursos.

Segundo Ball (2014), licenças são instrumentos eficazes para comunicar permissões sobre dados a potenciais reutilizadores. Licença é um instrumento legal para uma entidade detentora de direitos permitir que outra utilize uma obra original, sem que infrinja os direitos detidos. Sobre licenças e licenciamento, Ball (2014, p. 3) tece os seguintes comentários:

A primeira coisa a notar é que apenas o detentor dos direitos (ou alguém com um direito ou licença para agir em seu nome) pode conceder uma licença; portanto, é imperativo que os direitos de propriedade intelectual (DPI) relativos aos dados sejam estabelecidos antes de qualquer licenciamento. A segunda coisa a notar é que, embora seja da natureza de uma licença expandir em vez de restringir o que um licenciado pode fazer, algumas licenças são apresentadas dentro de contratos e os contratos podem colocar restrições adicionais ao licenciado e, de fato, ao licenciante (BALL, 2014, p. 3, *tradução nossa*).

Em geral, as licenças concedem permissões mediante o cumprimento de determinados termos. Embora sejam observadas variações em relação a considerações estabelecidas em licenças, três requisitos são comumente percebidos nesse contexto: atribuição, *copyleft* e não comercialidade. O primeiro requisito refere-se à devida atribuição de crédito pela obra ao licenciante, na medida em que essa for distribuída ou utilizada para derivação de nova obra. O segundo exige a liberação sob a mesma licença, e apenas essa licença, para a realização de qualquer trabalho derivado do licenciado. Finalmente, o terceiro requisito refere-se a uma licença não comercial que visa impedir que o licenciado explore a obra comercialmente (BALL, 2014).

Dentre os vários tipos de licenças que dizem respeito ao contexto eletrônico e a padrões para conjuntos de dados abertos, as licenças *Creative Commons* e *Open*

Data Commons são as mais utilizadas. A primeira cumpre o objetivo de estimular pesquisadores a definir as condições e termos os quais suas obras serão utilizadas. Já a segunda, refere-se à iniciativa promovida pela *Open Knowledge Foundation* (OKF), e abrange vários instrumentos jurídicos que auxiliam a criação e uso de dados abertos (SILVA, 2019).

Segundo Sayão e Sales (2015), alguns pesquisadores atribuem o conceito da licença *Creative Commons* para seus conjuntos de dados de pesquisa, e, dessa forma, explicitam os direitos mantidos e renunciados quando outros pesquisadores reutilizarem seus dados. Entretanto, tal licença não é apropriada para dados, havendo outras mais adequadas para essa finalidade, como é o caso da licença *Open Data Commons*.

Considerando a relevância das licenças *Creative Commons* e *Open Data Commons* como instrumentos que norteiam o reuso de dados em termos legais, a próxima seção dará ênfase à descrição de cada uma delas, e nessa ordem.

2.3.3.1 Creative Commons

Segundo informações coletadas no *site* da *Creative Commons*¹⁷, as licenças dessa espécie têm a mesma duração do prazo correspondente ao direito de autor ou aos direitos conexos. Licenciantes também podem adicionar outras autorizações em relação às condições nas quais o trabalho pode ser reutilizado. Assim sendo, as licenças *Creative Commons* são:

- **Domínio Público CC0** – permite que os licenciantes abram mão de todos os direitos sobre o trabalho, disponibilizando-o em domínio público;
- **Atribuição CC BY** – permite a distribuição, alteração, combinação e desenvolvimento do trabalho original, ainda que para fins comerciais, desde que haja a devida atribuição de autoria. É uma licença recomendada para aqueles que prezam a máxima divulgação e uso do conteúdo licenciado;

- **Atribuição-Compartilha Igual CC BY-SA** – permite a alteração, adaptação e desenvolvimento do trabalho original, ainda que para fins comerciais, desde que haja o devido reconhecimento de autoria e que novas criações derivadas sejam licenciadas sob as mesmas condições. Geralmente, essa licença é comparada a licenças relacionadas à *software* livre e de código aberto;
- **Atribuição-Sem Derivações CC BY-ND** – permite o reuso de obras para quaisquer finalidades, incluindo fins comerciais, desde que o trabalho original seja distribuído sem alterações e que se atribua devidamente os créditos ao autor;
- **Atribuição-Não Comercial CC BY-NC** – permite a alteração, adaptação e desenvolvimento do trabalho original para fins não comerciais. Embora novos trabalhos necessitem atribuir o devido crédito e não possam ser usados com propósitos comerciais, reutilizadores não precisam licenciar trabalhos derivados sob as mesmas condições;
- **Atribuição-Não Comercial-Compartilha Igual CC BY-NC-SA** – permite a alteração, adaptação e desenvolvimento do trabalho original para fins não comerciais, desde que haja o devido reconhecimento de autoria e que novas criações sejam licenciadas sob as mesmas condições;
- **Atribuição-Não Comercial-Sem Derivações CC BY-NC-ND** – sendo muito restritiva em relação às licenças apresentadas, essa só permite que outros realizem *download* dos trabalhos originais e os compartilhem mediante a devida atribuição de autoria, mas sem a possibilidade de alterá-los de forma alguma, ou utilizá-los para propósitos comerciais.

2.3.3.2 Open Data Commons

De acordo com informações coletadas no *site* do *Open Data Commons*¹⁸, há um conjunto de instrumentos legais e licenças que auxiliam pesquisadores na publicação e utilização de dados abertos. Assim sendo, as licenças *Open Data Commons* podem ser brevemente descritas da seguinte maneira:

18 Disponível em: <https://opendatacommons.org/>

- **Open Data Commons Open Database License (ODbL)**¹⁹ – é uma licença que permite aos usuários compartilhar, alterar e utilizar livremente bancos de dados, estendendo essa liberdade a outros. Geralmente, bancos de dados possuem conteúdo diversificado, incluindo imagens, materiais audiovisuais, entre outros recursos informacionais. Nesse sentido, destaca-se que a ODbL vigora somente no âmbito dos direitos sobre o banco de dados, e não de seu conteúdo, individualmente. Há ocasiões em que o conteúdo de um banco de dados é legalmente amparado por outros direitos, que podem incluir contratos particulares, marca registrada sobre determinado nome, ou direito de proteção sobre dados;
- **Open Data Commons Attribution License (ODC-By)**²⁰ – é uma licença que permite aos usuários compartilhar, alterar e utilizar livremente banco de dados, desde que sejam atendidos os devidos requisitos de atribuição quando esses bancos forem transmitidos publicamente ou quando forem criados bancos de dados derivados;
- **Open Data Commons Public Domain Dedication and License (PDDL)**²¹ – licença que permite o livre compartilhamento, alteração e utilização de trabalhos para qualquer fim e sem qualquer restrição. Destina-se ao uso de bancos de dados e de seu conteúdo, ou seja, abrange os dados em nível de coleções ou individualmente. Isso quer dizer que tanto banco de dados, como seu conteúdo (dados) podem ser utilizados para fins comerciais, sendo possível combiná-los e compartilhá-los, livremente, sem a necessidade de criação de cópia dessa licença, ou de atribuição ao criador de dados ou banco de dados. O objetivo a ser atingido é a eliminação de restrições relacionadas ao criação e utilização desses recursos informacionais.

Tão importante quanto a abertura de dados, é a própria dinâmica desses em relação à produção do conhecimento científico. Por essa razão, a próxima seção apresentará modelos do ciclo de vida de dados de pesquisa presentes na literatura

19 Disponível em: <https://opendatacommons.org/licenses/odbl/1-0/>

20 Disponível em: <https://opendatacommons.org/licenses/by/1-0/>

21 Disponível em: <https://opendatacommons.org/licenses/pddl/1-0/>

científica e técnica. Para tanto, foram examinados artigos científicos, relatórios técnicos, guias e páginas de instituições de ensino e pesquisa, como universidades e institutos de pesquisa.

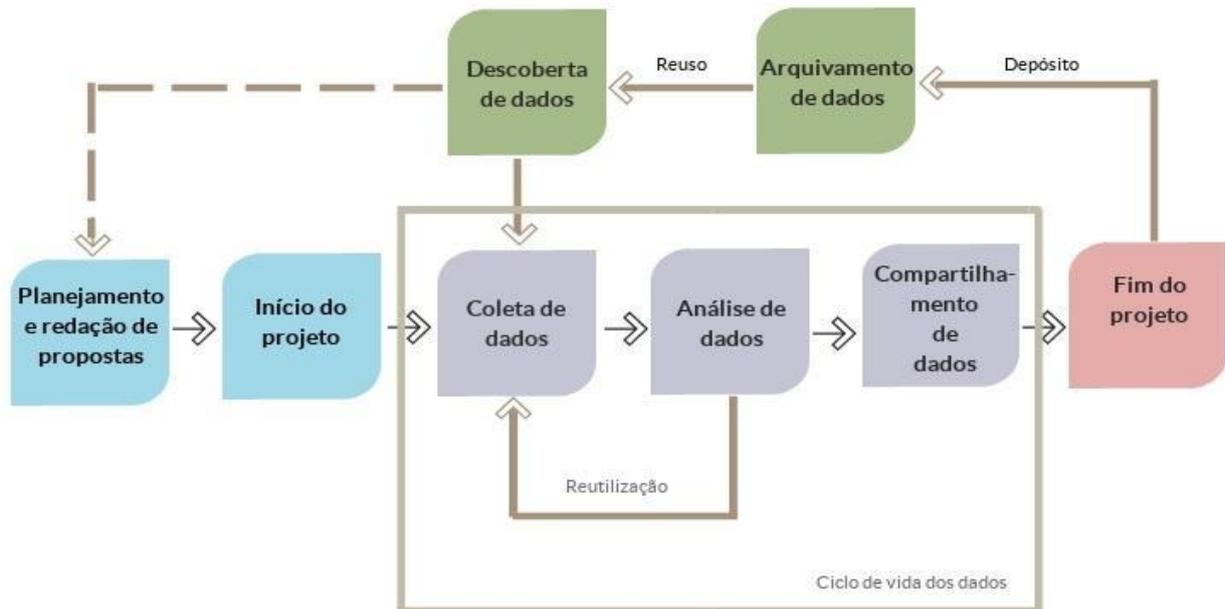
2.4 Modelos do ciclo de vida de dados de pesquisa

Com o propósito de ilustrar a variedade de conceitos, atores e processos envolvidos em representações do ciclo de vida de dados, serão comentados, nesta ordem, nove modelos elaborados por: Universidade de Virgínia; *Data Observation Network for Earth* (DataONE); *UK Data Archive*; *United States Geological Survey* (USGS); *Inter-university Consortium for Political and Social Research* (ICPSR); *German Federation for Biological Data* (GFBio); Jisc; biblioteca da *Technical University of Denmark* (DTU); Rüegg *et al.*

2.4.1 Modelo do ciclo de vida de dados de pesquisa formulado pela Universidade de Virgínia

A Figura 5 apresenta o modelo do ciclo de vida de dados elaborado pela Universidade de Virgínia²²:

Figura 5 – Modelo do ciclo de vida de dados de pesquisa formulado pela Universidade de Virginia



Fonte: traduzido e adaptado de Universidade de Virginia

Conforme demonstrado na Figura 5, as etapas que constituem o ciclo de vida de dados de pesquisa são:

- **planejamento e redação de propostas** – o pesquisador planeja aspectos importantes do projeto de pesquisa, como a coleta, disponibilidade e arquivamento de dados, bem como analisa custos e identifica os usuários potenciais dos dados de pesquisa obtidos;
- **início do projeto** – o pesquisador elabora um Plano de gestão de dados²³;
- **coleta de dados** – o pesquisador define padrões de qualidade para a coleta de dados e estabelece o controle de acesso aos dados, conforme critérios de segurança;
- **análise de dados** – o pesquisador analisa documentos e arquivos de dados;

23 Para fins deste trabalho, o termo “Plano de Gestão de dados” é compreendido tal como foi definido pela Organização das Nações Unidas para a Educação, a Ciência e a Cultura (UNESCO) na publicação intitulada *Guidelines for a Research Data Management Plan* (2016, p. 6, tradução nossa): Um plano de gestão de dados é um documento formal que descreve como lidar com os dados coletados ou gerados no decorrer de um projeto de pesquisa e o que acontece com esses dados durante seu ciclo de vida. O objetivo do plano de gestão de dados é garantir que os dados sejam coletados, documentados, disponibilizados e preservados adequadamente para uso futuro. (...) O plano de gestão de dados descreve os dados que serão coletados, as práticas de gerenciamento de dados que serão usadas, as políticas que serão aplicadas aos dados, quem terá a custódia dos dados, quem terá acesso aos dados e quem será responsável pela preservação dos dados.

- **compartilhamento de dados** – o pesquisador compartilha dados conforme requisitos definidos por instituições e agências financiadoras de pesquisas;
- **término do projeto** – armazenamento de dados em um repositório de dados de pesquisa.

2.4.2 Modelo do ciclo de vida de dados de pesquisa formulado pela DataOne

De acordo com o guia *Data management guide for public participation in scientific research*, publicado por grupo de trabalho da *DataOne*, em 2013, o ciclo de vida de dados acompanha toda a trajetória do projeto da pesquisa científica. À vista disso, a Figura 6 apresenta o modelo do ciclo de vida de dados formulado pela *DataOne*, no qual observam-se oito etapas: planejar; coletar; assegurar (qualidade); descrever; preservar; descobrir; integrar; analisar. Apesar de serem apresentadas conforme sequenciamento lógico, essas etapas podem ocorrer em sequência diferente, ao mesmo tempo, ou com o predomínio de uma etapa em detrimento de outra.

Figura 6 – Modelo do ciclo de vida de dados de pesquisa formulado pela DataOne



Fonte: traduzido e adaptado de DataONE (2013, p. 2)

Conforme apresentado na Figura 6, as etapas do ciclo de vida de dados de pesquisa são descritas da seguinte maneira:

- **planejar** – etapa iniciada quando os objetivos do projeto de pesquisa são definidos. O pesquisador mapeia os processos que serão realizados e os recursos a serem utilizados durante a pesquisa. Além disso, são elaborados um Plano de gestão de dados e planos de sustentabilidade de dados;
- **coletar** – etapa na qual os métodos mais apropriados são adotados na coleta, estruturação e organização dos dados de pesquisa;
- **assegurar (qualidade)** – emprego de padrões de qualidade nos procedimentos adotados em relação aos dados de pesquisa. Abrange a capacitação de participantes da pesquisa, a elaboração de instrumentos de avaliação dos dados, a identificação de vulnerabilidades e a correção de erros;
- **descrever** – descrição de dados de pesquisa, por meio de metadados (dados sobre dados), com o propósito de facilitar a sua recuperação, e, por conseguinte, seu compartilhamento e reuso. Como melhores práticas, recomenda-se a padronização de metadados, a documentação do processamento de dados e o armazenamento adequado dos mesmos;
- **preservar** – compreende o plano de preservação de dados a curto e longo prazo. Uma das mais conhecidas práticas de preservação de dados a curto prazo é a realização de *backups* manuais ou automáticos. Dentre as melhores práticas relacionadas à preservação a longo prazo, destaca-se a elaboração de política de *backup* de dados, o armazenamento de dados em formatos de arquivos estáveis, o depósito de dados em repositórios que adotem padrões de segurança e a identificação de dados sensíveis;
- **descobrir** – adoção de estratégias que agregam valor aos dados, atribuindo-lhes maior impacto e visibilidade a outros pesquisadores. Inclui atividades como o armazenamento de dados em repositórios de acesso aberto, o registro de dados em diretórios *online* de pesquisas e a utilização de metadados para descrevê-los;
- **integrar** – compreende a combinação de dados derivados de diversas fontes, possibilitando novas análises e o desenvolvimento de outras investigações;

- analisar – uso de *softwares* para auxiliar o processo de análise dos dados.

2.4.3 Modelo do ciclo de vida de dados de pesquisa formulado pelo UK Data Archive

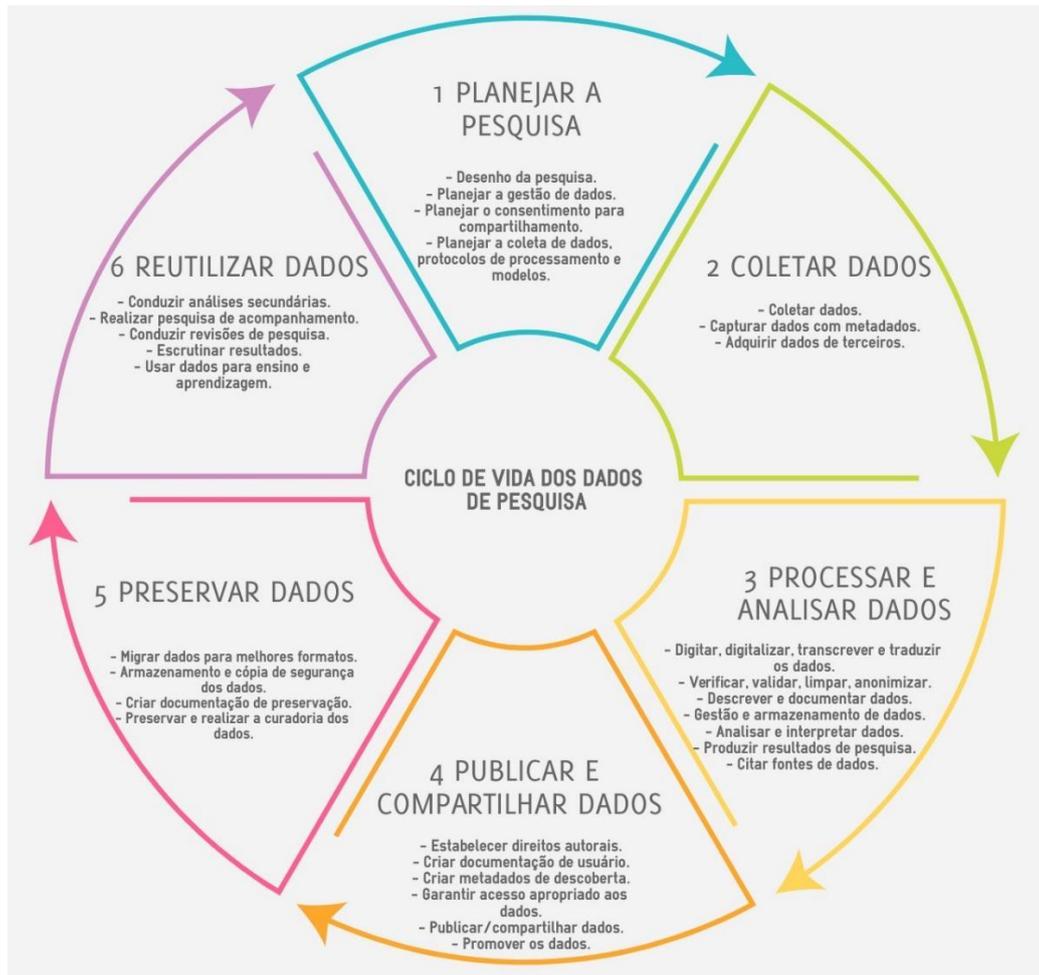
*UK Data Archive*²⁴ é um centro de referência internacional na aquisição, curadoria e acesso a dados de pesquisa nas áreas de Ciências Sociais e Humanidades. Fundado em 1967, na *University of Essex*, tem a finalidade de gerir dados de pesquisa de alta qualidade para que esses possam ser analisados e reutilizados. Conforme entendimento da instituição, dados de pesquisa podem ter uma vida útil superior aos projetos científicos que os geraram. Nessa direção, pesquisadores continuariam a usufruir desses dados, mesmo após o término do financiamento de uma pesquisa, podendo analisá-los e combiná-los a outros recursos informacionais. Por isso, quando dados de pesquisa são devidamente organizados, documentados, preservados e compartilhados, esses podem contribuir inestimavelmente para o avanço científico e para a aprendizagem e inovação.

A Figura 7 apresenta o modelo do ciclo de vida de dados de pesquisa proposto pelo *UK Data Archive*²⁵, constituído por seis etapas: 1) planejar a pesquisa; 2) coletar dados; 3) processar e analisar dados; 4) publicar e compartilhar dados; 5) preservar dados; 6) reutilizar dados.

24 Disponível em: <https://www.data-archive.ac.uk/about/>

25 Disponível em: <https://www.youtube.com/watch?v=-wjFMMQD3UA>

Figura 7 – Modelo do ciclo de vida de dados de pesquisa formulado pelo UK Data Archive



Fonte: traduzido e adaptado de UK Data Archive

Conforme observado na Figura 7, depende-se que a primeira etapa do ciclo de vida de dados – **planejar a pesquisa** – diz respeito a todo o planejamento da pesquisa em relação aos dados. Abrange a elaboração do Plano de gestão de dados e a definição de diretrizes e protocolos para a coleta, processamento e compartilhamento de dados. A segunda etapa – **coletar dados** – refere-se aos procedimentos de coleta de dados produzidos ou obtidos por pesquisadores. A terceira etapa – **processar e analisar dados** – compreende os procedimentos de tratamento técnico e análise de dados de pesquisa, incluindo atividades como a digitalização, a transcrição e a tradução de dados, a descrição, a anonimização, a validação e a interpretação de dados.

Depreende-se da Figura 7, ainda, que a quarta etapa do ciclo de vida de dados de pesquisa – **publicar e compartilhar dados** – refere-se à definição de requisitos para a publicação e compartilhamento de dados de pesquisa. Nessa etapa, determina-se como será realizada a indicação de autoria, as permissões de acesso aos dados e a criação de metadados. A quinta etapa – **preservar dados** – compila todos os esforços para que dados se mantenha íntegros, disponíveis e acessíveis ao longo do tempo. Em razão disso, abrange atividades como o armazenamento e cópias de segurança de dados, a elaboração de documentação de preservação de registros e a prática de curadoria de dados. Por fim, a sexta etapa – **reutilizar dados** – compreende a realização de estudos a partir de dados secundários, a revisão e apuração de resultados, e a aplicação de dados para o ensino.

2.4.4 Modelo do ciclo de vida de dados de pesquisa formulado pelo USGS

De acordo com relatório do *United States Geological Survey* (USGS), publicado em 2013, a gestão de dados de pesquisa é fundamental para a execução de um projeto científico. Além disso, modelos do ciclo de vida de dados constituem ferramentas visuais que auxiliam as práticas de pesquisadores em diferentes etapas da pesquisa. Isso posto, a Figura 8 apresenta o modelo elaborado por USGS que contribui para a compreensão de tais etapas.

Figura 8 – Modelo do ciclo de vida de dados de pesquisa formulado por USGS



Fonte: traduzido e adaptado de USGS (2013, p. 2)

Conforme pode ser visto na Figura 8, o modelo do ciclo de vida de dados de pesquisa elaborado por USGS é composto por seis elementos primários – planejar, adquirir, processar, analisar, preservar, publicar/compartilhar – representados por retângulos e três elementos transversais – descrever (metadados, documentação), gerenciar a qualidade, *backup* e segurança – representados por setas. Os elementos primários são descritos da seguinte maneira:

- **planejar** – etapa na qual pesquisadores arquitetam as atividades relacionadas à gestão de dados, desde a gênese do projeto de pesquisa até a publicação e arquivamento dos mesmos. Nessa direção, o Plano de gestão de dados é elaborado, as abordagens metodológicas são estabelecidas, os recursos necessários à condução da pesquisa são estimados e os resultados os quais se pretende atingir são definidos;
- **adquirir** – procedimentos pelos quais dados são coletados, gerados ou avaliados para o reuso. No âmbito do USGS, os dados incluem fluxogramas, mapas, registros biológicos, entre outros recursos informacionais;
- **processar** – compreende o processamento de dados coletados, o que inclui a combinação de diferentes conjuntos de dados, procedimentos de extração, transformação e carregamento de arquivos e preparação de dados para a análise;
- **analisar** – procedimentos relacionados à interpretação de dados processados. Abrange atividades como a elaboração de resumos, criação de gráficos, modelagem e análise estatística;
- **preservar** – refere-se à acessibilidade e ao armazenamento de dados, metadados e documentos complementares a longo prazo, de modo a assegurar seu reuso no futuro;
- **publicar/compartilhar** – compreende o compartilhamento de dados por meio da *World Wide Web*, catálogos de dados, mídias sociais e outros canais.

Apesar de os elementos primários remeterem-se a procedimentos específicos adotados em cada uma das etapas referidas, outros procedimentos estão presentes durante todas as etapas do ciclo de vida de dados de pesquisa, os quais foram

denominados elementos transversais. Por sua vez, esses são descritos da seguinte forma:

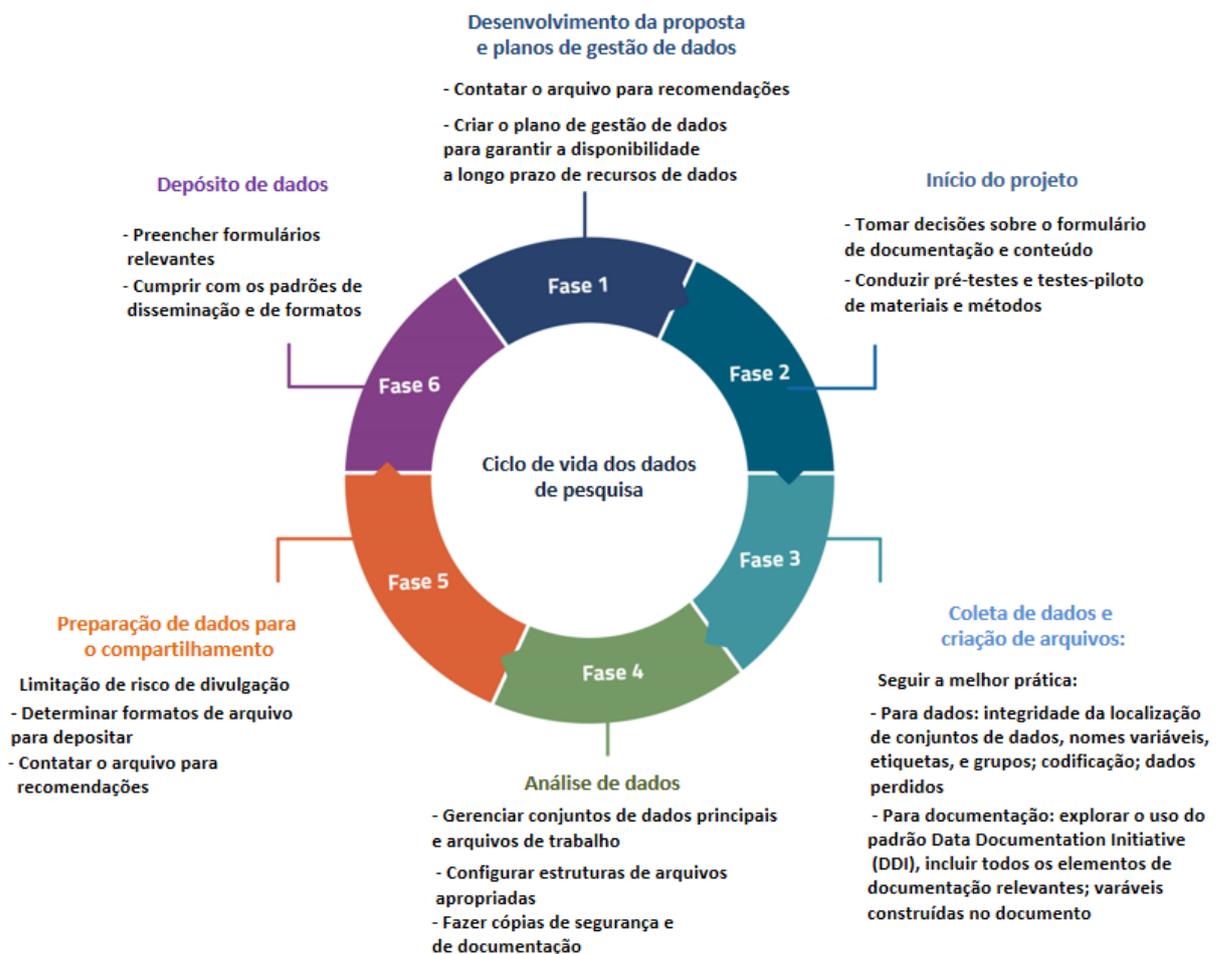
- **descrever (metadados, documentação)** – abrange a documentação minuciosa de rotinas de pesquisadores no decorrer de cada etapa do ciclo de vida de dados. Inclui o detalhamento do Plano de gestão de dados, de modo suficiente para que outros pesquisadores possam analisar, replicar e validar resultados da pesquisa. Nesse sentido, a utilização de metadados padronizados e a documentação em relação aos dados favorecem o entendimento do contexto de criação dos dados, a compilação, o processamento, a indexação e o reuso de dados de pesquisa;
- **gerenciar a qualidade** – elemento transversal que ocorre desde o início do projeto de pesquisa e estende-se às demais etapas do ciclo de vida de dados. Abrange o conjunto de procedimentos relacionados ao monitoramento contínuo dos dados de pesquisa;
- **Backup e segurança** – tendo em vista o acesso futuro aos dados, pesquisadores realizam *backups*, continuamente. Nesse sentido, evitam-se perdas de dados ocasionados, principalmente, por problemas técnicos em *hardwares* ou *softwares*, catástrofes naturais e falhas humanas. Além disso, medidas de proteção contra perdas de dados de pesquisa estendem-se a dados brutos ou processados, planos de gestão de dados, procedimentos de coleta e análise de dados, entre outras atividades.

2.4.5 Modelo do ciclo de vida de dados de pesquisa formulado pelo ICPSR

O *Inter-university Consortium for Political and Social Research* (ICPSR) foi firmado em 1962, e hoje tem se destacado no contexto de pesquisas na área de Ciências Sociais na realização de treinamentos sobre acesso, curadoria e procedimentos de análise de dados. Além disso, o arquivamento de dados viabilizado pelo ICPSR abrange dados de diversas disciplinas como Ciência Política, Sociologia, Economia, História, Educação, Saúde Pública, Demografia e Direito (ICPSR, 2012).

A Figura 9 apresenta o diagrama esquemático que representa o ciclo de vida de dados com ênfase no processo de arquivamento de dados, proposto pelo ICPSR, em 2012, na publicação intitulada: *Guide to social science data preparation and archiving: best practice throughout the data life cycle*. Esse diagrama estrutura-se mediante a ocorrência de seis fases: 1) desenvolvimento da proposta e planos de gestão de dados; 2) início do projeto; 3) coleta de dados e criação de arquivos; 4) análise dos dados; 5) preparação de dados para compartilhamento; 6) depósito dos dados. Todavia, é importante ressaltar que a realidade pode não ocorrer de modo tão linear como sugerido no diagrama.

Figura 9 – Modelo do ciclo de vida de dados de pesquisa formulado por ICPSR



Fonte: traduzido e adaptado de ICPSR (2012, p. 8)

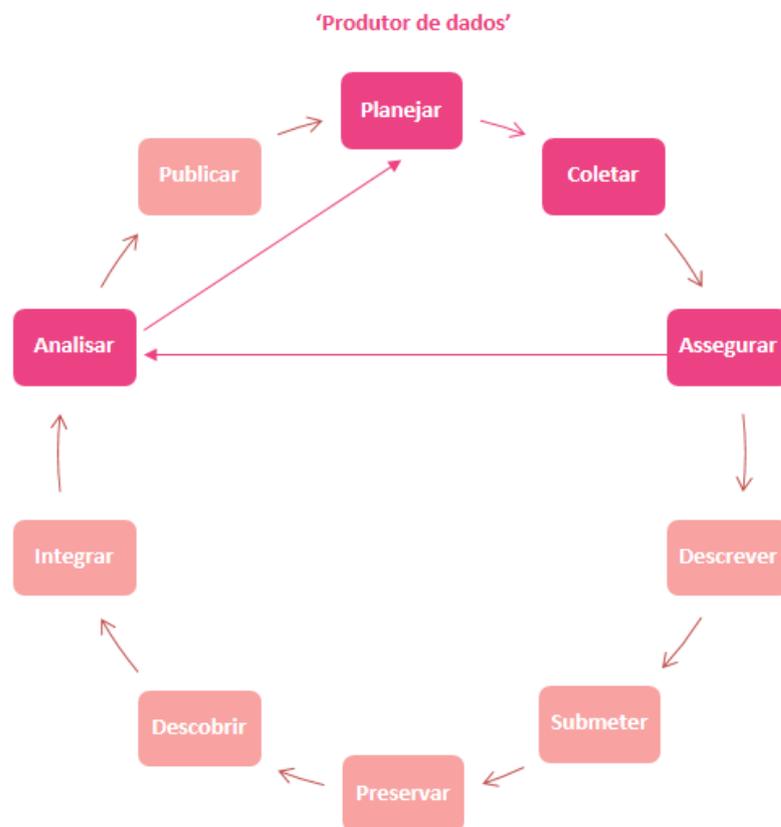
As fases que constituem o ciclo de vida de dados apresentado na Figura 9 são descritas da seguinte maneira (ICPSR, 2012):

- 1) **desenvolvimento da proposta e planos de gestão de dados** – compreende a elaboração do Plano de gestão de dados ainda nos estágios iniciais da proposta de pesquisa. Esse plano pode ser um instrumento muito útil para antever problemas relacionados à realização de estudos;
- 2) **início do projeto** – após o recebimento de financiamento, o projeto de pesquisa tem início. Nessa fase, o pesquisador estima quais recursos são necessários à execução do projeto, estuda como reduzir significativamente custos, além de refletir sobre questões estruturais à pesquisa, como a nomeação de arquivos, a captura, a organização, a integração, o armazenamento de dados e a utilização de metadados;
- 3) **coleta de dados e criação de arquivos** – refere-se a práticas estratégicas para a coleta de dados de pesquisa, diante da diversidade de tipos de dados e de formatos disponíveis;
- 4) **análise de dados** – fase em que dados de pesquisa são alterados mediante acréscimos ou reduções;
- 5) **preparação de dados para o compartilhamento** – abrange os procedimentos que viabilizam o compartilhamento de dados, mediante observação a níveis de acesso preestabelecidos e à confidencialidade de participantes de pesquisas;
- 6) **depósito de dados** – fase na qual dados são armazenados conforme o atendimento de requisitos e padrões, como o modelo de referência *Open Archival Information System* (OAIS). Nessa fase, pesquisadores atentam-se à preservação digital de dados, viabilizando o seu arquivamento e legibilidade por diferentes *softwares*.

2.4.6 Modelo do ciclo de vida de dados de pesquisa formulado pela GFBio

German Federation for Biological Data (GFBio)²⁶ é uma iniciativa que se propõe a facilitar o compartilhamento de dados no âmbito das Ciências Biológicas e ambientais e a apoiar pesquisadores na implementação de estratégias efetivas para a gestão de dados de pesquisa. A GFBio auxilia todas as etapas do ciclo de vida de dados, desde a proposta do projeto, até o arquivamento de dados a longo prazo. A Figura 10 apresenta o ciclo de vida de dados de pesquisa idealizado pela GFBio, como uma ferramenta conceitual que contribui para o entendimento das diferentes etapas que o constitui. No entanto, esse ciclo pode ocorrer de modo não semelhante, considerando a diversidade de contextos nos quais pesquisas são realizadas e a execução parcial de alguns procedimentos no decorrer do ciclo.

Figura 10 – Modelo do ciclo de vida de dados de pesquisa formulado pela GFBio



Fonte: traduzido e adaptado de GFBio

Como mostrado na Figura 10, o modelo do ciclo de vida de dados de pesquisa elaborado pela GFBio²⁷ é composto por dez etapas: 1) planejar; 2) coletar; 3) assegurar (qualidade); 4) descrever; 5) submeter; 6) preservar; 7) descobrir; 8) integrar; 9) analisar; 10) publicar. Tais etapas são detalhadas da seguinte maneira:

- 1) **planejar** – etapa na qual o pesquisador elabora o Plano de gestão de dados de pesquisa, preferencialmente, na fase de proposta da pesquisa. Refere-se ao planejamento sobre a coleta de dados, métodos a serem adotados, requisitos para o compartilhamento e reuso de dados, e previsão de recursos tecnológicos necessários;
- 2) **coletar** – fase em que o pesquisador captura dados, geralmente, por meio da utilização de recursos tecnológicos, como gravadores, computadores e satélites. Dados de pesquisa podem ser produzidos ou obtidos a partir de observações, entrevistas, medições, simulações por algoritmos, literatura científica, entre outras fontes;
- 3) **assegurar (qualidade)** – refere-se ao monitoramento contínuo da qualidade de dados durante todo o seu ciclo de vida. A qualidade dos dados é um dos principais fatores estimados para o seu reuso. Nesse sentido, dados devem ser acessíveis, precisos, legíveis e interpretáveis;
- 4) **descrever** – utilização de metadados com o propósito de conjuntos de dados serem facilmente localizados e reutilizados por outros usuários. Metadados são úteis para esclarecer o contexto de produção de dados, de modo a responder questões como o porquê de os dados terem sido gerados, quem os produziu/obteve, onde e quando foram coletados, quais conteúdos abordam e como foram avaliados;
- 5) **submeter** – refere-se à transferência de dados do domínio particular do pesquisador para o domínio de pesquisa compartilhado. Uma vez que a submissão de conjuntos de dados para um centro de dados que assegure sua preservação a longo prazo não implica, automaticamente, no acesso aberto

aos dados, é necessário que o pesquisador indique quais termos e licenças de reuso e citação de dados serão adotados;

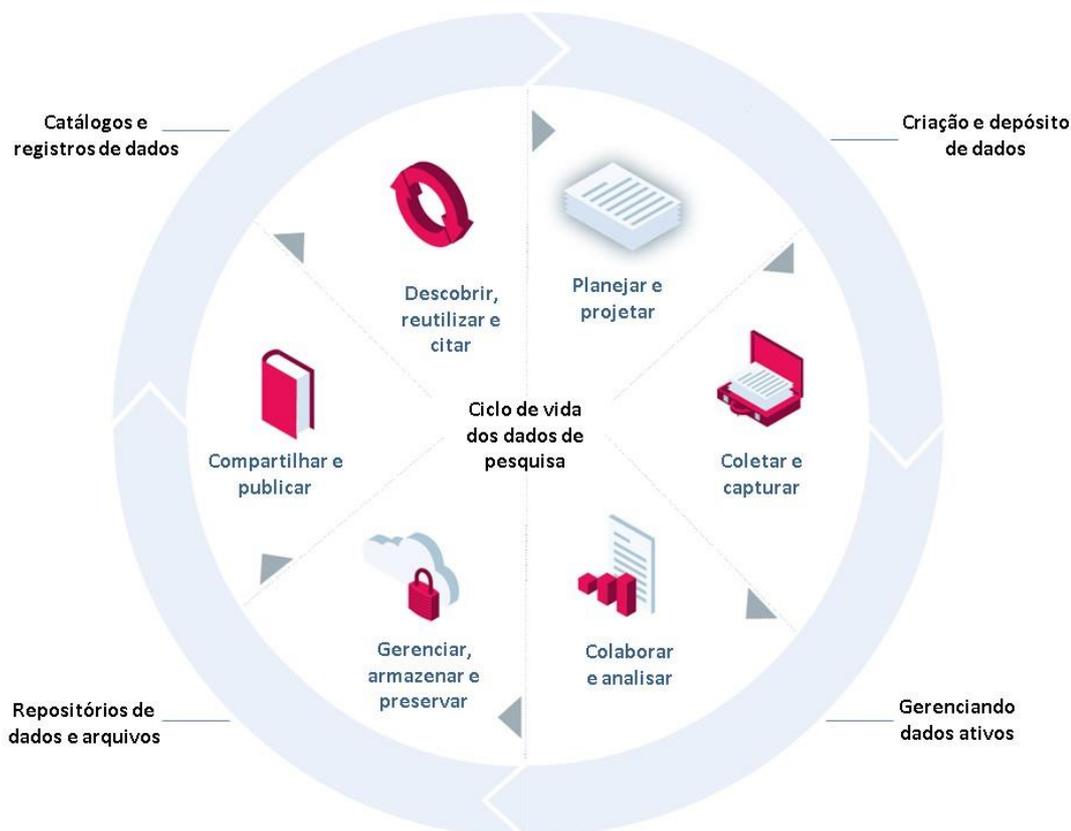
- 6) **preservar** – compreende procedimentos que assegurem a integridade de dados a longo prazo, mantendo-os acessíveis (recuperáveis e utilizáveis), autênticos (não manipuláveis ou alteráveis) e reutilizáveis. Nesse sentido, cópias de segurança são apenas componentes da preservação. Além disso, atualmente, observam-se diferentes métodos de preservação digital, como migração, emulação e arqueologia digital.
- 7) **descobrir** – etapa na qual pesquisadores descobrem dados disponíveis, o que pode ser muito valioso para o desenvolvimento de seu projeto de pesquisa. Para tanto, é importante que dados e metadados estejam visíveis, acessíveis e possuam um registro de identificador persistente;
- 8) **integrar** – refere-se à integração e à combinação de dados, resultando em um conjunto de dados ainda maior. O pesquisador compara resultados, testa novas hipóteses e reutiliza dados mediante atribuição de autoria;
- 9) **analisar** – compreende a análise de conjuntos de dados, de modo a obter uma nova visão sobre questões investigadas e outras conclusões. Para tanto, *softwares* podem ser utilizados.
- 10) **publicar** – etapa na qual o pesquisador publica os seus dados. Com frequência, editores de periódicos científicos têm requisitado a disponibilização de dados para a publicação de artigos. Isso devido o entendimento de que dados contribuem para a transparência da pesquisa, validação de resultados e reprodutibilidade de estudos. A publicação de dados também ocorre em periódicos de dados.

2.4.7 Modelo do ciclo de vida de dados de pesquisa formulado pela Jisc

De acordo com a Jisc – organização sem fins lucrativos atuante no campo de tecnologias digitais para educação e pesquisa no Reino Unido – a gestão de dados de pesquisa, em nível institucional, é uma tarefa complexa. Com o propósito de

auxiliar pesquisadores nessa tarefa, a organização propôs um modelo do ciclo de vida de dados de pesquisa²⁸, conforme pode ser observado na Figura 11.

Figura 11 – Modelo do ciclo de vida de dados de pesquisa formulado pela Jisc



Fonte: traduzido e adaptado de Jisc

Como apresentado na Figura 11, o modelo do ciclo de vida de dados de pesquisa elaborado pela Jisc estrutura-se, sobretudo, em seis etapas: 1) planejar e desenhar; 2) coletar e capturar; 3) colaborar e analisar; 4) gerenciar, armazenar e preservar; 5) compartilhar e publicar; 6) descobrir, reutilizar e citar. Essas etapas são detalhadas do seguinte modo:

- 1) **planejar e projetar** – considerando variáveis como políticas, serviços e requisitos da instituição a qual vincula-se o pesquisador, essa etapa refere-se ao planejamento da gestão de dados, antes mesmo do início da pesquisa. O planejamento contribui para que pesquisadores organizem, armazenem e

preservem seus dados em conformidade com os princípios FAIR, além de sinalizar os recursos necessários à pesquisa e os principais desafios a serem enfrentados;

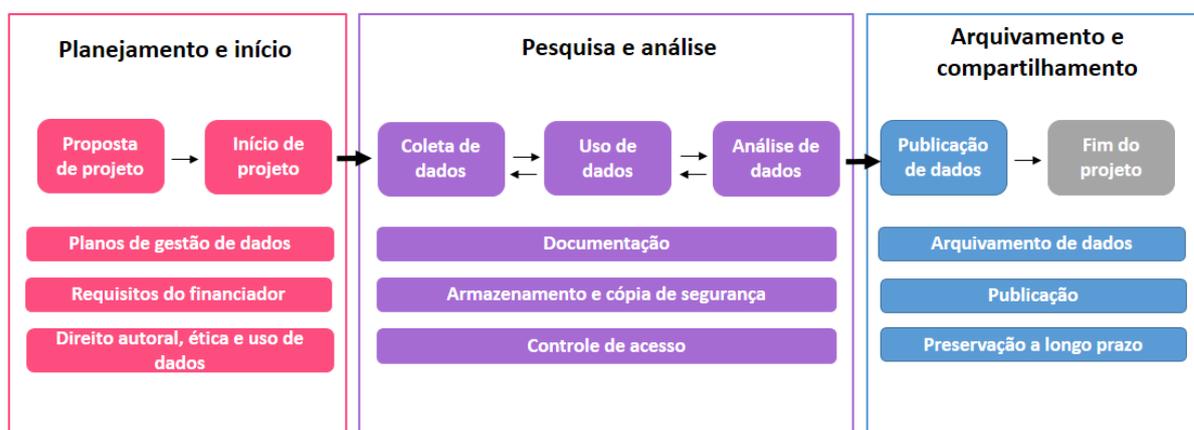
- 2) **coletar e capturar** – coletam-se dados, de modo a organizá-los e a documentá-los. O pesquisador documenta os dados de pesquisa, adicionando-lhes metadados. Esses, por seu turno, contêm informações sobre dados, tornando-os localizáveis, contextualizáveis, reutilizáveis e citáveis. Nessa etapa, o pesquisador ainda padroniza formatos de arquivo, nomeia e estrutura pastas, uma vez que tais formatos afetarão a compatibilidade com outros arquivos, e, conseqüentemente, o reuso de dados;
- 3) **colaborar e analisar** – considerando a colaboração entre pesquisadores em nível nacional, ou internacional, essa etapa refere-se à documentação de processos de gestão de dados. Além disso, o compartilhamento de arquivos é uma atividade central no âmbito de projetos colaborativos;
- 4) **gerenciar, armazenar e preservar** – etapa na qual o pesquisador adota medidas para a proteção de dados de pesquisa, o que inclui considerações técnicas e legais a respeito do armazenamento e preservação de conjuntos de dados;
- 5) **compartilhar e publicar** – fase em que o pesquisador compartilha dados de várias maneiras, seja informalmente durante um projeto, por meio de *e-mails* ou serviços de compartilhamento de arquivos, seja formalmente no fim de um projeto, por meio de repositórios de dados, *data journals* e materiais suplementares de publicações. Ao compartilhar e publicar dados, o pesquisador dá oportunidades para que outros possam descobrir seu trabalho mais facilmente, podendo esse ser citado em publicações, devido ao uso de identificadores persistentes nos conjuntos de dados. Também é nessa etapa que o pesquisador define os níveis de acesso que terão os seus dados. Isso porque a escolha de uma licença inadequada pode ocasionar a inutilização de todo um trabalho de pesquisa. Além disso, a publicação de dados é uma tarefa complexa, que abrange o contexto de normas disciplinares, de infraestrutura técnica, de treinamentos, entre outros aspectos;

- 6) **descobrir, reutilizar e citar** – etapa integrada à perspectiva de usuários de dados de pesquisa, de modo que esses possam encontrar dados ao visitarem repositórios disciplinares, gerais ou institucionais, a partir de *links* disponíveis em artigos, ou até mesmo solicitando o reuso de conjunto de dados diretamente com o pesquisador que os produziu. É interessante notar que o reuso de dados de pesquisa deve ocorrer em consonância com termos, licenças e condições associados aos dados, e com os direitos de propriedade intelectual estabelecido por autores.

2.4.8 Modelo do ciclo de vida de dados de pesquisa formulado pela biblioteca da DTU

Segundo a biblioteca da *Technical University of Denmark* (DTU)²⁹, a gestão de dados de pesquisa não se restringe ao arquivamento de dados em um disco rígido. Ao contrário, é uma prática relevante ao longo do projeto de pesquisa, incluindo o planejamento, a coleta de dados, a publicação e o arquivamento em repositórios. A Figura 12 apresenta aspectos da gestão de dados de pesquisa e sua relação com as etapas do ciclo de vida de dados. No entanto, é interessante notar que raramente os processos de pesquisa são tão lineares como são ilustrados nesse modelo.

Figura 12 – Modelo do ciclo de vida de dados de pesquisa formulado pela biblioteca da DTU



Fonte: traduzido e adaptado de DTU

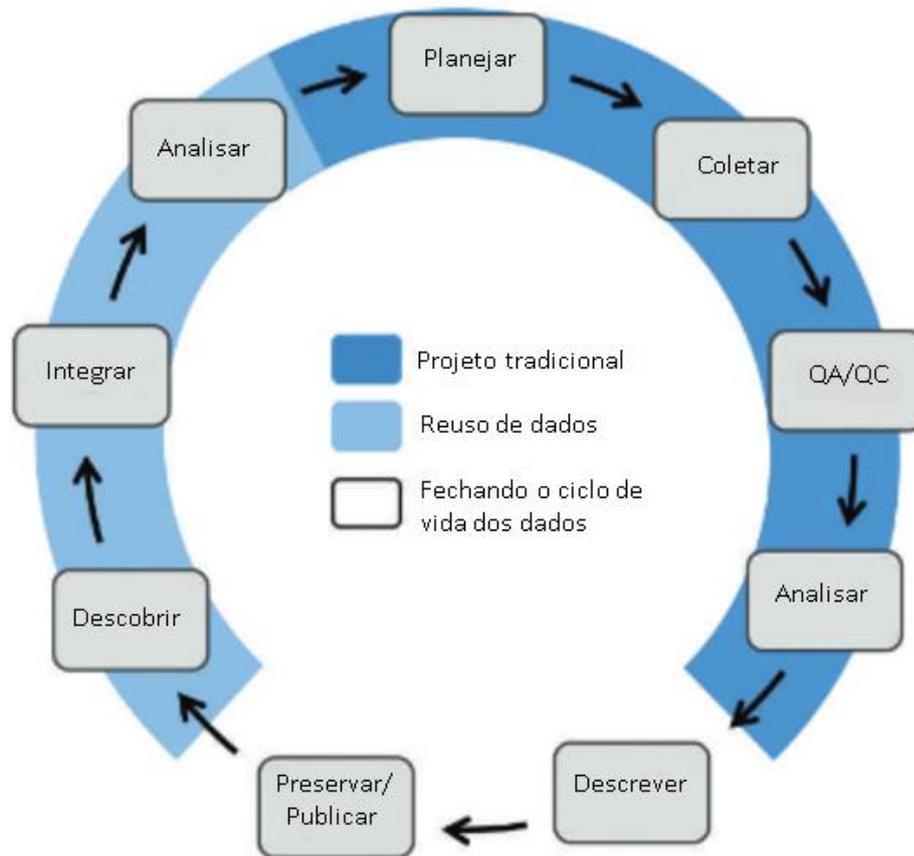
Conforme apresentado na Figura 12, as principais etapas que constituem o modelo do ciclo de vida de dados de pesquisa elaborado pela biblioteca da DTU são o planejamento e início da pesquisa, a pesquisa e análise, e o arquivamento e compartilhamento. Os procedimentos relativos a tais etapas são detalhados nos seguintes tópicos:

- **planejamento e início da pesquisa** – etapa fundamental no processo de tomada de decisão que influenciará o modo pelo qual dados de pesquisa serão gerenciados, organizados e preservados. Para tanto, o pesquisador elabora o Plano de gestão de dados, atualizando-o conforme o andamento do projeto de pesquisa e de acordo com requisitos estabelecidos por financiadores da pesquisa;
- **pesquisa e análise** – etapa relativa à estruturação/organização da coleta de dados e às primeiras análises na pesquisa. Para tanto, o pesquisador descreve detalhadamente como os dados serão coletados, organizados e analisados conforme a adoção de protocolos e padrões, bem como realiza cópias de segurança;
- **arquivamento e compartilhamento** – considerando que a gestão de dados de pesquisa não termina quando um projeto de pesquisa é concluído, ou quando o financiamento à pesquisa é finalizado, nessa etapa o pesquisador documenta e armazena seus dados.

2.4.9 Modelo do ciclo de vida de dados de pesquisa formulado por Rüegg et al.

De acordo com Rüegg *et al.* (2014), projetos de pesquisas na área de ecologia de macrossistemas demandam práticas de gestão de dados desde o seu início, de modo a propiciar um ambiente científico mais aberto. Isso porque tais projetos geralmente abordam problemas complexos, que requerem coleta de dados de fontes variadas e a colaboração interdisciplinar entre pesquisadores. A Figura 13 apresenta o ciclo de vida de dados elaborado pelos autores, que ilustra um projeto de pesquisa tradicional.

Figura 13 – Modelo do ciclo de vida de dados de pesquisa formulado por Rüegg *et al.*



Fonte: traduzido e adaptado de Rüegg *et al.* (2014)

Conforme observado na Figura 13, Rüegg *et al.* (2014) descrevem um projeto de pesquisa tradicional (representado em azul escuro), no qual há o planejamento, a coleta de dados, o controle de qualidade, e a análise de dados. As siglas (QA/QC), referem-se, respectivamente, aos termos em inglês: *quality assurance* e *quality control*. Em português, ambos estão relacionados à obtenção de qualidade de dados coletados.

Geralmente, esse ciclo é concluído quando o pesquisador armazena conjuntos de dados em seu computador pessoal, com pouca ou nenhuma documentação sobre os mesmos. Conseqüentemente, muitos dados acabam sendo perdidos, ao passo que instrumentos necessários para uma gestão eficiente são desproporcionais diante do “dilúvio de dados” existentes, da complexidade dos problemas científicos e das diferenças disciplinares.

Assim, muitos pesquisadores não recebem treinamento para a gestão de dados de pesquisa, denotando uma reflexão tardia (e não financiada) a respeito da importância do tema. Em razão disso, Rüegg *et al.* (2014) destacam duas etapas cabíveis de serem adotadas, com vistas à disponibilidade a longo prazo e ao reuso de dados de pesquisa, sendo elas: “Descrever / documentar” e “Preservar / publicar”. Para tanto, é preciso que dados estejam devidamente vinculados a metadados que os descrevam, e, em seguida, sejam arquivados de modo permanente, possibilitando o seu reuso além da vida de um projeto específico. Outrossim, a documentação e a publicação de dados de pesquisa são fundamentais para facilitar a pesquisa em escalas espaciais e temporais no âmbito de pesquisas em ecologia de macrossistemas, e permitem o desenvolvimento das etapas “Descobrir”, “Integrar” e “Analisar” em outros projetos (representado em azul claro na Figura 13).

Ante aos estudos referidos, é possível afirmar que o entendimento sobre o ciclo de vida de dados é peça-chave para promover a gestão de dados no âmbito de pesquisas científicas. Por essa razão, a seção seguinte discutirá a gestão de dados de pesquisa sob diversas perspectivas, a partir da literatura científica e técnica.

2.5 Gestão de dados de pesquisa

Wilkinson *et al.* (2016) afirmam que uma “boa gestão” é fundamental para promover o reuso de dados e a consequente integração do conhecimento científico. Apesar de não haver consenso que defina o que constitui uma “boa gestão”, os autores ressaltam que muitos dados são subutilizados, e, em razão disso, são necessárias práticas que vão muito além da coleta adequada e arquivamento de dados. Nesse sentido, é preciso reconhecer a noção de curadoria de dados a longo prazo, viabilizando a descoberta e reuso dos mesmos.

Considerando o contexto exposto, é válido destacar a iniciativa dos princípios FAIR como um instrumento norteador que pode ser referência na realização de práticas institucionais de gestão de dados de pesquisa. De acordo com Wilkinson *et al.* (2016), no ano de 2014, foi realizado em Leiden, Holanda, um *workshop* chamado *Jointly Designing a Data Fairport*, que reuniu acadêmicos e *stakeholders* interessados

na superação de barreiras relacionadas à descoberta e ao reuso de dados de pesquisa. No *workshop* foi discutida a necessidade da elaboração de princípios fundamentais que orientem práticas em prol da consecução de dados facilmente descobertos, acessíveis, disponíveis para o reuso e apropriadamente citados. À vista disso, um conjunto denominado FAIR – composto por quatro princípios: *findable*, *accessible*, *interoperable* e *reusable* – foi formulado nesse encontro. A sigla traduzida para o português refere-se, respectivamente, a características de dados localizáveis, acessíveis, interoperáveis e reutilizáveis. O Quadro 3 apresenta esses princípios:

Quadro 3 – Princípios do FAIR

LOCALIZÁVEIS
<ul style="list-style-type: none"> 1 (Meta) dados são atribuídos a um identificador globalmente único e persistente 2 Dados são descritos com metadados ricos 3 Metadados incluem de forma clara e explícita o identificador dos dados que descreve 4 (Meta) dados são registrados ou indexados em um recurso pesquisável
ACESSÍVEIS
<ul style="list-style-type: none"> 1 (Meta) dados são recuperáveis pelo seu identificador usando um protocolo de comunicação padronizado <ul style="list-style-type: none"> 1.1 O protocolo é aberto, gratuito e universalmente implementável 1.2 O protocolo permite um procedimento de autenticação e autorização, quando necessário 2 Metadados estão acessíveis, mesmo quando os dados não estão mais disponíveis
INTEROPERÁVEIS
<ul style="list-style-type: none"> 1 (Meta) dados usam uma linguagem formal, acessível, compartilhada e amplamente aplicável para a representação do conhecimento. 2 (Meta) dados usam vocabulários que seguem os princípios do FAIR 3 (Meta) dados incluem referências qualificadas a outros (meta) dados
REUTILIZÁVEIS
<ul style="list-style-type: none"> 1 Meta (dados) são ricamente descritos com uma pluralidade de atributos precisos e relevantes <ul style="list-style-type: none"> 1.1. (Meta) dados são liberados com uma licença de uso de dados clara e acessível 1.2. (Meta) dados estão associados à proveniência detalhada 1.3. (Meta) dados atendem aos padrões da comunidade relevantes ao domínio

Fonte: traduzido e adaptado de Wilkinson *et al.* (2016, p. 4)

É interessante notar que Wilkinson *et al.* (2016) ressaltam que tais princípios se aplicam não somente a dados de pesquisa no sentido tradicional, mas estende-se ao contexto de outros tipos de dados, como algoritmos, ferramentas e fluxos de trabalho. Além disso, os autores destacam que diversos *stakeholders* estão envolvidos e podem se beneficiar com a adoção dos princípios FAIR, como: os pesquisadores que almejam compartilhar seus dados e reutilizar dados de outros pesquisadores, os criadores de *softwares* e ferramentas que auxiliam o tratamento e análise de dados,

as agências públicas e privadas de fomento à pesquisa, que tem se preocupado em relação à preservação de dados a longo prazo, e a comunidade de dados científicos, que tem estudado novas maneiras de mineração, integração e análise para descobrir dados.

Wilkinson *et al.* (2016) ainda relatam que princípios FAIR têm sido implementados em várias plataformas de dados em âmbito internacional, os quais destacam-se: *FAIRDOM*, *Open PHACTS*, *wwPDB* e *UniProt*. Outrossim, a *Ligue des Bibliothèques Européennes de Recherche* (LIBER) ressalta que bibliotecas deveriam se sentir confiantes na aplicação desses princípios, uma vez que são instituições com conhecimento, *expertise* e longa tradição na descrição, disponibilização e preservação a longo prazo de recursos informacionais dispostos em meio impresso ou digital. Em razão disso, a LIBER elencou algumas ações que podem ser implementadas em bibliotecas em atendimento ao FAIR, as quais se destacam:

- promover os princípios do FAIR para as equipes de pesquisa local e de tecnologia da informação;
- inserir os princípios do FAIR em seus planos de gestão de dados e em suas práticas e políticas de preservação digital;
- oferecer treinamentos a bibliotecários sobre metadados, vocabulários e ferramentas inerentes ao contexto de dados de pesquisa;
- estimular pesquisadores a depositar seus dados em arquivos, conforme as recomendações dos princípios do FAIR;
- avaliar em sua instituição como tem sido realizada a gestão de dados de pesquisa em relação aos princípios do FAIR.

Com base no ciclo de vida de dados, Patel (2016) propôs uma estrutura conceitual para a gestão de dados de pesquisa em instituições, considerando as etapas de criação, armazenamento, organização e compartilhamento de dados. Nesse contexto, o autor destacou a importância de discutir aspectos relacionados principalmente aos direitos autorais, às licenças dos dados, às interpretações errôneas sobre dados, à segurança e à privacidade dos dados e à mudança de atitude de pesquisadores a favor do compartilhamento e reuso de dados. O Quadro 4 apresenta a estrutura conceitual para a gestão de dados de pesquisa elaborada por Patel

(2016). Nele, são identificadas três camadas principais: gestão de dados de pesquisa; armazenamento e hospedagem de dados; e uso de dados; sob as quais depreendem-se algumas subcamadas.

Quadro 4 – Estrutura conceitual para gestão de dados de pesquisa

CAMADA 1: GESTÃO DE DADOS
<ul style="list-style-type: none"> - Desenvolvimento de política institucional para o compartilhamento de dados - Mudança de mentalidade de pesquisadores - Coleção de dados para pesquisadores - Direitos autorais e licenças de dados - Referência cruzada de dados para metodologia - Classificação de dados - Anonimização de dados - Descrição e identificação de dados - Organização de dados - Estrutura de interoperabilidade para dados
CAMADA 2: ARMAZENAMENTO E HOSPEDAGEM DE DADOS
<ul style="list-style-type: none"> - Seleção de formatos de arquivos - Geração de dados por meio de colaborações entre instituições públicas e privadas - Serviço de hospedagem de dados - Contribuidores independentes de dados - Responsabilidade com dados hospedados - Segurança dos dados - Software de hospedagem de dados - Cópia de segurança dos dados
CAMADA 3: USO DE DADOS
<ul style="list-style-type: none"> - Acesso aos dados - Direito autoral e licença de dados - Direitos de trabalhos derivados de dados

Fonte: traduzido e adaptado de Patel (2016, p. 229)

No tocante à camada gestão de dados, destacam-se as seguintes subcamadas (PATEL, 2016):

- **desenvolvimento de política institucional para o compartilhamento de dados** – promove a cultura mandatória do compartilhamento de dados em instituições. É necessário que tal política estabeleça com clareza seu propósito e abranja orientações sobre temas como a criação de metadados, direitos autorais, condições para a disponibilização de dados, entre outros;

- **mudança de mentalidade de pesquisadores** – é preciso motivar pesquisadores quanto ao compartilhamento de dados abertos, tendo em vista os possíveis benefícios que essa prática pode trazer ao campo científico e à sociedade em geral;
- **coleta de dados para pesquisadores** – recomenda-se que conjuntos de dados sejam integrados com a publicação de resultados de pesquisas, documentando-os em seus diversos estágios;
- **direitos autorais e licença de dados** – é fundamental que projetos de pesquisas estabeleçam, inequivocamente, os termos para a concessão de licenças de uso de dados, em consonância com os direitos autorais;
- **referência cruzada de dados para metodologia** – é recomendável que repositórios de dados incluam mecanismos que façam referência cruzada de dados à metodologia utilizada em sua coleta, durante a pesquisa;
- **classificação de dados** – processo que possibilita a identificação de dados compartilháveis, ou com algum tipo de restrição de disponibilização, e que facilita a recuperação de dados;
- **anonimização de dados** – é preciso assegurar métodos efetivos para a anonimização;
- **descrição e identificação dos dados** – processos que favorecem a recuperação de dados, logo, é essencial o uso de padrões de metadados;
- **organização de dados** – atualmente têm sido desenvolvidas ontologias para a organização de dados, de modo a torná-los mais significativos e recuperáveis aos usuários;
- **estrutura de interoperabilidade para dados** – deve ser estabelecida, uma vez que dados de diferentes disciplinas podem ser coletados em formatos variáveis. Com relação à camada de armazenamento e hospedagem de dados, ressaltam-se as seguintes subcamadas (PATEL, 2016):
 - **seleção de formatos de arquivos** – processo estrutural para armazenamento de dados, de modo a garantir a acessibilidade aos dados a longo prazo;
 - **geração de dados por meio de colaborações entre instituições públicas e privadas;**

- **serviço de hospedagem de dados** – recomenda-se que instituições disponham de seus próprios repositórios de dados para hospedar dados;
- **contribuidores independentes de dados** – na gestão de dados de pesquisa também é importante considerar pesquisadores que não possuem vínculo com nenhuma instituição;
- **responsabilidade com dados hospedados;**
- **segurança dos dados;**
- **software de hospedagem dos dados;**
- **cópia de segurança dos dados.**

Quanto à camada de uso de dados, são apresentadas as seguintes subcamadas (PATEL, 2016):

- **acesso aos dados** – decisão na qual se define a abertura ou restrição de dados;
- **direito autoral e licença dos dados;**
- **direitos de trabalhos derivados de dados.**

De acordo com Yu (2017), políticas mandatórias para a disseminação de dados e para a elaboração de planos de gestão de dados são cada vez mais requeridas por agências de fomento à pesquisa e por instituições governamentais. Conseqüentemente, isso tem impulsionado o aumento de serviços de gestão de dados em universidades, em especial, no contexto de bibliotecas. Tais serviços abrangem todas as etapas do ciclo de vida de dados de pesquisa, ou seja, incluem o plano de gestão de dados, a coleta, a análise e o armazenamento de dados, a documentação da pesquisa, a criação de metadados, o arquivamento, a descoberta e o compartilhamento de dados.

Sendo assim, na conhecida era do *Big Data*, a provisão de serviços de dados de pesquisa tem se tornado uma das funções fundamentais de bibliotecas. Muitos são os desafios atrelados a esse contexto, porém, também há diversas oportunidades para pesquisadores e bibliotecários, como: explorar as necessidades relativas à gestão de dados de pesquisa em instituições de ensino e pesquisa; redimensionar o fornecimento de serviços em razão do ciclo de vida de dados de pesquisa; desenvolver infraestrutura tecnológica de suporte a dados; promover o depósito de dados em

repositórios de acesso aberto entre pesquisadores; garantir a preservação de dados a longo prazo (YU, 2017).

Segundo Sales *et al.* (2019), a participação de bibliotecários na gestão de dados gerados por pesquisadores ocorre em várias fases do processo da pesquisa. Antes de essa ser iniciada, bibliotecários podem auxiliar pesquisadores em seu planejamento e na elaboração do Plano de gestão de dados. Durante a pesquisa, tais profissionais podem prestar assistência a pesquisadores na etapa de coleta de dados e na seleção de repositórios apropriados à publicação dos dados produzidos. Concluída a pesquisa, podem auxiliar pesquisadores na publicação e no armazenamento de dados, de forma segura e a longo prazo.

Fruto de pesquisa cujo objetivo foi estudar e sistematizar a função dos bibliotecários e as competências necessárias à atuação na gestão de dados de pesquisa, Sales *et al.* (2019) identificaram competências fundamentais àqueles que almejam trabalhar nessa atividade. Sendo assim, o Quadro 5 apresenta tais competências, distribuídas conforme as fases da gestão de pesquisa:

Quadro 5 – Competências dos bibliotecários na gestão de dados de pesquisa

ANTES DA PESQUISA	DURANTE A PESQUISA	DEPOIS DA PESQUISA
Auxiliar pesquisadores na elaboração do plano de gestão de dados	Tipificar dados de pesquisa	Auxiliar na publicação de dados (identificação de repositórios ou outras formas de publicação ex: data journal, periódico de resultado negativo, etc)
Planejar a curadoria	Conhecer a estrutura informacional do dado de pesquisa e o seu ciclo de vida	Auxiliar na contextualização, isto é, na documentação de conjuntos de dados (definições, metodologia de coleta, etc)
Identificar fluxos de trabalhos (ou mais especificamente o fluxo da pesquisa)	Administrar o ciclo de vida dos dados de pesquisa, desde sua geração/coleta, bem como seleção e desenvolvimento de coleção.	
Identificar recursos e infraestruturas para manutenção e promoção de dados de pesquisa	Organizar dados de pesquisa / atribuir metadados gerais e disciplinares	
Orientar a organização de arquivos de dados e o uso de ferramentas de gestão de dados de pesquisa		
Apoiar na identificação e escolha de ferramentas adequadas para análise, processamento e visualização	Apoiar a análise de dados e o processamento, indicando ferramentas e promovendo treinamentos	Apoiar a visualização de dados, indicando ferramentas e provendo treinamentos
Apoiar a adoção de práticas de gestão de dados de pesquisa em parceria com departamentos, grupos de pesquisa, comissões, etc.;		
Entender e promover preservação digital, isto é, elaboração de política de preservação	Entender e promover preservação digital, isto é, gerenciamento de versões, armazenamento e backup	Entender e promover preservação digital de longo prazo, isto é, desenvolvimento de ambientes confiáveis para preservação
Contribuir para a elaboração de políticas institucionais de dados de pesquisa	Gerenciar sistemas de armazenamento de dados	Promover o reuso de dados, através de divulgação e seleção de dados adequados
Conhecer aspectos legais dos dados de pesquisa, bem como as leis de direitos autorais	Promover a capacitação para o desenvolvimento da competência em gestão de dados de pesquisa (research data literacy)	Auxiliar na elaboração de citação e referência de dados
Definir políticas de acesso	Criar e oferecer tutoriais sobre a elaboração de planos de gestão de dados	

Fonte: Adaptado de Sales *et al.* (2019, p. 311)

Por meio da análise do Quadro 5, Sales *et al.* (2019) demonstram que diversas atividades podem ser desempenhadas no âmbito da gestão de dados de pesquisa. Algumas delas requererem capacitação específica por parte de bibliotecários. Enquanto outras, assemelham-se a atividades tradicionalmente desenvolvidas por bibliotecas e centros de informação. Além disso, há atividades que percorrerão todas as fases da gestão de dados. Por essa razão, demanda-se que bibliotecários desenvolvam habilidades específicas e que demonstrem conhecimentos sobre tópicos relacionados a políticas de gestão de dados, à curadoria de dados, a *softwares* de armazenamento de dados, e à capacitação de usuários.

Tendo como base os argumentos de Yu (2017) e de Sales *et al.* (2019) que destacam a importância da mediação do bibliotecário no contexto da produção científica, as seções seguintes dedicar-se-ão à discussão acerca do papel exercido por esse profissional na gestão de dados de pesquisa, e de serviços de gestão de dados de pesquisa em bibliotecas, nessa ordem.

2.5.1 O papel do bibliotecário de dados na gestão de dados de pesquisa

No âmbito desta pesquisa, bibliotecários de dados podem ser entendidos na perspectiva de Chiware (2020), como sendo aqueles profissionais que trabalham com a gestão de dados de pesquisa. Utilizam dados como recursos informacionais e auxiliam pesquisadores no decorrer das atividades relacionadas à produção científica. Para tanto, é importante que esses profissionais estejam aptos tecnicamente a desenvolver soluções de gestão, arquivamento, compartilhamento, mineração de dados, e que participem de projetos de criação de repositórios institucionais que relacionem documentos publicados a conjuntos de dados.

Em estudo que objetivou descrever conhecimentos e habilidades essenciais à função do bibliotecário de dados, Federer (2018) teve como base a perspectiva de profissionais da informação que declararam estar exercendo tal função. Os resultados do estudo apontaram que bibliotecários de dados compõem uma comunidade heterogênea de profissionais da informação oriundos de contextos profissionais e educacionais diversificados, e que realizam diferentes práticas em relação à gestão

de dados de pesquisa. Em se tratando de uma profissão emergente, as opiniões dos bibliotecários de dados divergiram no que se refere às especializações necessárias para a ocupação do cargo e às práticas realizadas nesse contexto.

Além disso, dentre os participantes da pesquisa, observou-se a existência de dois grupos de profissionais: bibliotecários especialistas em um assunto e bibliotecários de dados generalistas. O primeiro grupo é integrado por profissionais que desenvolveram profundamente habilidades em determinada área do conhecimento e adquiriram vasta experiência trabalhando com usuários especialistas. Já o segundo grupo, refere-se àqueles profissionais que lidam com dados de pesquisa oriundos de várias áreas do conhecimento, não somente de um campo específico. Tais profissionais podem adequar-se facilmente em ambientes acadêmicos, visto que há oportunidades de realização de trabalhos colaborativos junto a professores, estudantes e pesquisadores em diferentes disciplinas (FEDERER, 2018).

Assim sendo, é importante que bibliotecas e instituições que solicitem a contratação de um bibliotecário de dados considerem as necessidades de seus usuários para verificar se essas podem ser mais bem atendidas por um bibliotecário especialista em um assunto, ou por um generalista (FEDERER, 2018).

Conforme o entendimento de Semeler e Pinto (2019), é fundamental que o bibliotecário de dados lance mão de suas habilidades tradicionais para o oferecimento de produtos e serviços compatíveis a um contexto caracterizado, sobretudo, pela produção de grandes quantidades de conjuntos de dados. Assim, é importante que esse profissional se comprometa em investigar sistemas e regulamentos relacionados a repositórios de dados de pesquisa, em aprimorar seus conhecimentos sobre coleta, tratamento técnico, análise e visualização de dados, e em desenvolver habilidades necessárias para trabalhar com diferentes tipos de dados.

Mesmo não sendo um programador, necessariamente, recomenda-se que um bibliotecário de dados compreenda aspectos relativos a linguagens e à lógica de programação de computadores, na medida em que lida com diversos *softwares* relacionados à gestão de dados e manuseia diferentes ferramentas de recuperação de informações em bases de dados (SEMELER; PINTO; ROZADOS, 2019).

Da mesma forma, é altamente desejável que esse profissional trabalhe com repositórios de dados e auxilie pesquisadores em várias fases da investigação científica, com vistas à adequada documentação, preservação e reutilização de dados. Ademais, é importante que o mesmo se interesse em participar de discussões que abordem tópicos como o compartilhamento de dados, o uso de ferramentas de acesso a dados e de formatos de metadados, e melhores práticas de preservação de dados (SEMELER; PINTO, 2020).

Para Martínez-Uribe e Fernández (2015), as habilidades essenciais a serem desenvolvidas pelo bibliotecário de dados incluem o manuseio de dados e metadados, conhecimentos na área de Informática, experiências em serviços de referência e capacitação de usuários. A essas habilidades somam-se o conhecimento de políticas de dados, a aptidão na comunicação com pesquisadores em diversas áreas e o domínio de ferramentas de gestão de recursos informacionais disponíveis em meio digital.

Considerando a importância do bibliotecário de dados e das bibliotecas durante o ciclo de vida de dados de pesquisa, a próxima seção abordará aspectos referentes à atuação dessas instituições no desenvolvimento de serviços de gestão de dados de pesquisa.

2.5.2 Serviços de gestão de dados de pesquisa em bibliotecas

Conforme entendem Martínez-Uribe e Fernández (2015), o serviço de dados em bibliotecas tem a função de assegurar qualidade a dados desde a sua criação, permitindo sua futura utilização e reuso. Para tanto, são imprescindíveis a criação de repositórios de dados e o desenvolvimento de infraestruturas que possibilitem o seu armazenamento, compartilhamento e publicação.

Em se tratando do ambiente acadêmico, Barbrow, Brush e Goldman (2017) destacam que a realização de pesquisas científicas tem sido responsável pela geração de grandes quantidades de dados, os quais não devem ser apenas processados e analisados por pesquisadores, mas também gerenciados no decorrer de todo o ciclo da pesquisa científica. Por essa razão, bibliotecas acadêmicas começam a oferecer

serviços de gestão de dados de pesquisa voltados à comunidade de usuários que atendem. Nessa direção, tais bibliotecas dão apoio ao corpo docente e a pesquisadores na elaboração do plano de gestão de dados (frequentemente requerido por agências de financiamento à pesquisa), bem como no arquivamento e compartilhamento de dados.

De modo semelhante, Thomas e Urban (2018) afirmam que serviços de dados estão crescendo no ambiente de bibliotecas acadêmicas, possibilitando uma gestão efetiva de dados de pesquisa. Os motivos que contribuem para explicar tal crescimento são: a intensa produção de dados; mandatos de gestão de dados expedidos por agências de financiamento à pesquisa; o fortalecimento de redes de compartilhamento de dados.

Segundo Rice (2019), cada vez mais as bibliotecas têm dado suporte à gestão de dados de pesquisa por meio de serviços de dados. Esses, por seu turno, incluem a realização de várias atividades que podem ser assim elencadas: criação de site dedicado ao serviço de gestão de dados de pesquisa; assistência a pesquisadores na elaboração de planos de gestão de dados; realização de treinamentos relacionados à gestão de dados; divulgação de melhores práticas relativas à curadoria de dados; realização de reuniões e eventos sobre gestão de dados de pesquisa.

Silva (2016) destaca que bibliotecas estão desenvolvendo serviços de gestão de dados durante as etapas que constituem o ciclo de vida de dados de pesquisa, significando a possibilidade de oferecer apoio a pesquisadores já na fase de produção de dados. Geralmente, os serviços prestados ocorrem mediante estreita colaboração entre bibliotecários e pesquisadores. Sob esse ponto de vista, há várias ações as quais o bibliotecário pode auxiliar a comunidade científica no desenvolvimento de seus projetos, principalmente no tocante à obtenção e preservação de dados.

No que diz respeito à etapa da obtenção de dados, o bibliotecário pode auxiliar o pesquisador na elaboração do Plano de gestão de dados, na utilização de metadados e na escolha de repositórios para armazenamento de dados. Após serem coletados, os dados passarão por processo de tratamento técnico realizado pelo bibliotecário, sendo transcritos, digitalizados e armazenados em repositórios apropriados. Com relação à etapa da preservação de dados, essa inclui atividades

de migração de dados para formatos atualizados e a realização de *backups*. Nessa direção, bibliotecários podem orientar pesquisadores quanto a requisitos necessários à preservação de dados, instruí-los sobre formatos de preservação, ou até mesmo realizar o procedimento de migração (SILVA, 2016).

De acordo com Semeler e Pinto (2020), dois tipos de serviços de dados são realizados em bibliotecas: serviços de curadoria de dados e serviços de gestão de dados. Os primeiros concentram seus esforços nas práticas de armazenamento seguro e preservação de dados a longo prazo. Já os segundos, auxiliam pesquisadores a lidar com os dados gerados no decorrer da pesquisa, o que inclui apoio na elaboração do Plano de gestão de dados, na organização e no armazenamento de dados.

De acordo com Martínez-Urbe e Fernández (2015), o trabalho de curadoria de dados consiste em gerenciar dados digitais ao longo de seu ciclo de vida, de modo a assegurar-lhes acesso presente e futuro. Para tanto, é necessário realizar procedimentos relacionados à codificação, limpeza, organização e normalização de conteúdo.

Semelhantemente, Noonan e Chute (2014) definem o termo curadoria de dados sob a perspectiva das diferentes etapas que constituem o ciclo de vida de dados, como se pode observar em seus comentários:

Curadoria de dados é a atividade de gestão de dados ao longo de seu ciclo de vida; apropriadamente mantendo sua integridade e autenticidade; garantindo que seja devidamente avaliado, selecionado, armazenado de forma segura e acessível; e oferecendo suporte a sua usabilidade em ambientes tecnológicos subsequentes (NOONAN; CHUTE, 2014, p. 203, *tradução nossa*).

Curadores de dados trabalham junto a pesquisadores, tratando de dados com diferentes formatos e padrões de metadados. Logo, é imprescindível que tais curadores demonstrem conhecimentos no campo da preservação digital (principalmente em sistemas de repositórios) e relacionados a ontologias e taxonomias aplicadas ao contexto de dados, bem como estejam aptos para analisar estruturas e os relacionamentos existentes entre dados (MARTÍNEZ-URIBE; FERNÁNDEZ, 2015).

Para Sales *et al.* (2019) a curadoria dos dados científicos é entendida como um serviço, que também é oferecido por bibliotecas, cuja finalidade é garantir acesso

ininterrupto e o reuso de dados ao longo do tempo. Nessa direção, é altamente relevante que as necessidades de pesquisadores em relação a dados sejam atendidas efetivamente, sendo respeitados a dimensão cultural e os modos pelos quais operam cada área do conhecimento.

Quanto ao ambiente digital, Sayão e Sales (2012) comentam sobre práticas da “curadoria digital”, que por seu turno, garantem a sustentabilidade de dados para o futuro, facilitando o acesso permanente aos mesmos. Além disso, a curadoria digital ascende como uma área com grandes oportunidades de diálogo entre diferentes disciplinas e categorias profissionais, conforme explicam os autores:

A curadoria digital, em resumo, assegura a sustentabilidade dos dados para o futuro, não deixando, entretanto, de conferir valor imediato a eles para os seus criadores e para os seus usuários. Os recursos estratégicos, metodológicos e as tecnologias envolvidas nas práticas da curadoria digital facilitam o acesso persistente a dados digitais confiáveis por meio da melhoria da qualidade desses dados, do seu contexto de pesquisa e da checagem de autenticidade. [...] A curadoria digital emerge como uma nova área de práticas e de pesquisa de espectro amplo que dialoga com várias disciplinas e muitos gêneros de profissionais (SAYÃO; SALES, 2012, p. 185).

No que concerne à gestão de dados de pesquisa, Silva (2016) explica que esse é um processo vivo de extrema importância principalmente a pesquisadores. Isso porque uma gestão de dados adequada contribui para que pesquisadores não percam tempo tentando localizar, organizar, entender ou (re)utilizar os próprios dados que produzem. Nesse sentido, reitera-se que “os dados devem trabalhar para o pesquisador e não contra ele; a pesquisa científica já é por si mesma suficientemente difícil” (SILVA, 2016, p. 404).

Com base nos estudos referidos, pode-se compreender que repositórios de dados de pesquisa são instrumentos fundamentais à gestão de dados, possibilitando o seu depósito, compartilhamento e preservação a longo prazo. Além disso, são meios que oferecem visibilidade a dados no âmbito da Internet. Pela importância que assumem nesse contexto, a próxima seção trata, especificamente, sobre repositórios de dados de pesquisa.

2.6 Repositório de dados de pesquisa

De acordo com Pinfield (2009, p. 165), um repositório pode ser entendido como um “conjunto de sistemas e serviços que facilitam a inclusão, o armazenamento, a gestão, a recuperação, a exibição e reuso de objetos digitais”. Repositórios podem disponibilizar acesso a vários tipos de objetos digitais, de modo a incluir artigos de periódicos, capítulos de livros, teses e dissertações, recursos de aprendizagens e conjuntos de dados.

Segundo informações no *site* do Serviço de Documentação da Universidade do Minho (SDUM)³⁰, o êxito de um repositório está relacionado a diversos aspectos, dentre os quais podem ser destacados: 1) a divulgação do repositório; 2) a existência de serviços de suporte; 3) a disseminação e ampliação da visibilidade de conteúdo; 4) a interoperabilidade com diferentes sistemas; 5) estar em conformidade com políticas de metadados e preservação.

Página na *Internet* do *National Institutes of Health (NIH)* – Estados Unidos da América³¹ – define repositório de dados como um local que deposita, organiza e disponibiliza dados de modo lógico. Repositórios de dados podem ter requisitos específicos relacionados a determinada área disciplinar, ao acesso a dados e seu reuso, aos formatos de arquivos e à estrutura de metadados. Alguns repositórios estabelecem restrições àqueles que almejam depositar seus dados, como nível de qualificação acadêmica e qualidade de dados. Contudo, outros são classificados como “abertos”, e, por conseguinte, se propõem a eliminar barreiras relativas à submissão e ao acesso a dados de pesquisa.

Segundo Uzwyshyn (2016), os repositórios de dados de pesquisa *online* são banco de dados robustos criados para o gerenciamento, compartilhamento, disponibilização e armazenamento de conjuntos de dados de pesquisa. Por meio deles, é possível disseminar, agregar, sintetizar, revisar e validar resultados de estudos. Além disso, possibilitam a publicação dos chamados “dados negativos”, que são aqueles gerados em experimentos mal-sucedidos. Isso permite que outros

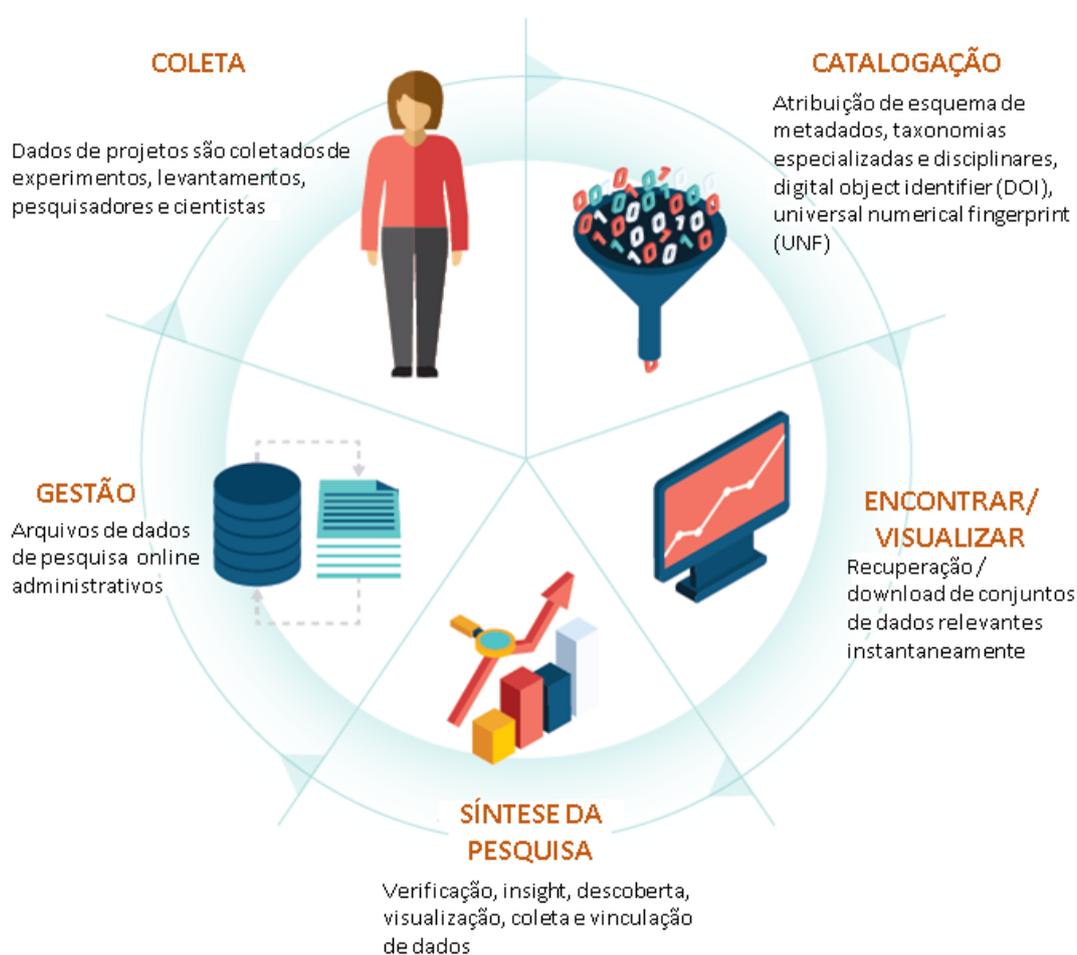
30 Disponível em: https://openaccess.sdum.uminho.pt/?page_id=348

31 Disponível em: <https://nlnm.gov/data/thesaurus/data-repository>

pesquisadores adotem procedimentos metodológicos diferentes, evita a duplicação de esforços e otimiza a utilização de recursos.

Do mesmo modo, tais repositórios asseguram a preservação de dados ao longo do tempo e, conseqüentemente, seu acesso continuado por meio de *links* permanentes. Além disso, a trajetória de infraestrutura de um repositório de dados ocorre mediante etapas de um ciclo de vida, como demonstrado na Figura 14 (UZWYSHYN, 2016):

Figura 14 – Ciclo de vida de um repositório de dados de pesquisa



Fonte: Traduzido e adaptado de Uzwysyn (2016, p. 19)

Conforme demonstrado na Figura 14 elaborada por Uzwysyn (2016), o ciclo de vida de um repositório de dados é iniciado quando o pesquisador coleta dados para o seu projeto. Posteriormente, segue em direção ao depósito e à catalogação de dados, processo no qual são atribuídos esquemas de metadados e taxonomias,

além do *Digital Object Identifier* (DOI) e *Universal Numerical Fingerprint* (UNF). Possibilitando acesso imediato a dados de pesquisa, é possível recuperá-los e realizar *downloads*. Conseqüentemente, isso permitirá a síntese de conjuntos de dados, análise e verificação de resultados de pesquisas. Por seu turno, a gestão de dados refere-se à manipulação de arquivos de dados de pesquisa *online* administrativos.

Para Rodrigues *et al.* (2010, p. 22), o termo “repositório” refere-se a um “sistema informático em que existe uma plataforma de armazenamento de objectos representados em ficheiros, capaz de incorporar novos objectos à medida que são produzidos ou submetidos”. Com relação aos repositórios de dados, essa noção de repositório de objetos se amplia, na medida em que cada conjunto de dados apresenta características intrínsecas e, em razão disso, é necessário que recebam tratamento diferenciado.

Levando em conta as especificidades relativas ao contexto dos dados de pesquisa, as ações de curadoria podem ocorrer em três níveis: 1) armazenamento; 2) representação do conjunto; 3) representação do item. O primeiro nível assegura que conjuntos de dados sejam depositados e armazenados seguramente, de modo que sejam preservados. O segundo nível, trata da descrição de conjuntos de dados por meio de metadados. Finalmente, o último nível abrange aspectos relacionados à descrição de itens de conjuntos de dados, o que inclui a caracterização individualizada de dados (RODRIGUES *et al.*, 2010).

Considerando o exposto em relação a aspectos que caracterizam os repositórios de dados de pesquisa, a próxima seção continuará o tópico discutido, abordando algumas de suas funções e benefícios à comunicação científica.

2.6.1 Funções e benefícios dos repositórios de dados de pesquisa

De acordo com Sayão e Sales (2016), repositórios de dados são sistemas digitais que integram várias funções para oferecer apoio aos processos referentes à gestão de dados de pesquisa. Em geral, tais funções ocorrem com mais frequência no nível administrativo e biblioteconômico, do que no nível tecnológico. Além disso, é ideal que tais repositórios sejam dirigidos por uma política institucional de gestão

de dados de pesquisa formalizada em documento de domínio público. Com base nisso, A Figura 15 ilustra um fluxo de gestão de dados de pesquisa operacionalizado em um repositório genérico, conforme diretrizes estabelecidas em uma política institucional de gestão de dados.

Figura 15 – Fluxo de gestão de dados de pesquisa



Fonte: Sayão e Sales (2016)

De maneira complementar, o Quadro 6 detalha possíveis funções que podem ser realizadas em cada etapa apresentada na Figura 15, que são: captura de dados; catalogação das coleções de dados; arquivamento e preservação; interoperabilidade; recuperação, acesso e reuso (SAYÃO, SALES, 2016):

Quadro 6 – Funções de um sistema de gestão de dados de pesquisa

<p>CAPTURA DE DADOS</p> <p>Inserção de coleções de dados – primários ou derivados - provenientes de experimentos, simulações, observações, questionários, levantamentos, etc. Os dados podem ser incorporados ao repositório pelos próprios autores - por autossucessão - ou por equipes especializadas vinculadas ao serviço.</p>	
<p>FUNÇÕES</p>	<ul style="list-style-type: none"> • Seleção dos dados passíveis de serem arquivados; • Verificação do enquadramento no escopo do repositório; • Verificação dos formatos de arquivos aceitáveis para submissão; • Verificação dos direitos associados às coleções (copyright e licenças); • Verificação de dados sensíveis (dados não anonimizados, confidenciais, pessoais); • Verificação do volume e quantidade de arquivos; • Verificação dos metadados gerais e disciplinares que acompanham os dados; • Normalização para elenco de formatos padronizados aceitos para arquivamento e disseminação; • Controle de qualidade dos dados; • Definição de tempo de embargo
<p>CATALOGAÇÃO DAS COLEÇÕES DE DADOS</p> <p>Descrição, atribuição de metadados e inclusão de documentação que assegurem que os dados possam ser acessados e interpretados no tempo e no espaço.</p>	
<p>FUNÇÕES</p>	<ul style="list-style-type: none"> • Atribuição de: metadados descritivos, estruturais, administrativos, técnicos (que inclui os relativos às dependências técnicas dos objetos digitais) • Atribuição de metadados de preservação, que assegurem a proveniência, autenticidade e integridade dos dados ao longo do tempo; • Uso de taxonomias especializadas e disciplinares; • Atribuição de identificador persistente (DOI, Handles, UNF, URN, etc.) que permita que os dados possam ser localizados de forma persistente e citados como as publicações acadêmicas; • Identificação do autor (ORCID ID, Scopus Author ID, ResearcherID, etc.) • Inclusão de documentação sobre os dados, incluindo descrição do projeto, dos arquivos e dos parâmetros; cadernos de laboratório e de campo, protocolos de pesquisa ou metodologia, etc.; • Vinculação (por links) a publicações e a dados relacionados internos e externos ao repositório
<p>ARQUIVAMENTO E PRESERVAÇÃO</p> <p>Arquivamento seguro que garanta a gestão de curto e longo prazo das coleções de dados orientados por um plano/política de preservação digital.</p>	
<p>FUNÇÕES</p>	<ul style="list-style-type: none"> • Armazenamento em sistemas seguros; • Gestão da preservação de curto prazo (backups, backups redundantes offsite; checagem de integridade, armazenamento seguro, criptografia, compressão); • Gestão de longo prazo (migração, emulação, reformatação para formatos padronizados, aplicação de normas pertinentes (OAIS, TRAC), informação de fixidade voltada para validar a autenticidade e integridade de um objeto digital (checksums, assinatura digital) • Implementação de trilhas de auditoria.
<p>INTEROPERABILIDADE</p> <p>Intercâmbio e compartilhamento e linkage com outros repositórios de dados e outros sistemas de informação (repositórios institucionais, bibliotecas digitais de publicações acadêmicas, editoras científicas)</p>	
<p>FUNÇÕES</p>	<ul style="list-style-type: none"> • Disponibilização de metadados segundo o protocolo OAI-PMH; • Agregação para formação de publicações ampliadas segundo o padrão OAI-ORE; • Uso dos padrões, web service, linked data e outros; • Empacotamento de metadados para intercâmbio segundo o padrão METS;
<p>RECUPERAÇÃO, ACESSO E REUSO</p> <p>Interface web para a descoberta, acesso e downloading de coleções de dados relevantes para o usuário ou para aplicações computacionais, como visualização e mapeamento, que podem prover serviços a partir dessas coleções; vinculado a uma política de acesso estabelecida pela instituição que inclui: tempo de embargo, direito de acesso, pagamentos, restrições sobre determinadas coleções, acesso somente aos metadados; registros de usuários e termos de uso dos dados.</p>	
<p>FUNÇÕES</p>	<ul style="list-style-type: none"> • Disponibilização de interfaces web para recuperação, acesso e downloading; • Oferta de aplicações e serviços sobre as coleções;

Fonte: Sayão e Sales (2016, p. 100)

Conforme o entendimento de Sayão e Sales (2016), as ações desenvolvidas por repositórios, no sentido da gestão de dados de pesquisa, resultam em benefícios para todo o contexto da produção científica. Dentre esses, podem ser destacados os seguintes:

- **visibilidade dos dados** – considerando que apenas uma pequena parcela de dados de uma pesquisa costuma ser disponibilizada por meio de canais formais, repositórios expandem a visibilidade de dados, possibilitando que esses sejam localizados e citados com mais frequência;
- **compartilhamento de dados** – devido à capacidade de agregação e disponibilização de recursos informacionais, e por se constituírem meio de socialização entre comunidades de pesquisadores, repositórios são altamente relevantes para a promoção do compartilhamento de dados;
- **crédito ao autor dos dados** – repositórios permitem identificar a autoria de coleções de dados de modo inequívoco e persistente. Assim, todo o fruto do trabalho intelectual de autores pode ser reconhecido, citado, avaliado e recompensado;
- **preservação digital** – proporcionam ambiente tecnológico adequado para a preservação de dados de pesquisa a longo prazo;
- **memória científica e transparência** – colaboram para a formação da memória científica em relação a dados produzidos em instituições e contribuem com os princípios de transparência aplicados à pesquisa;
- **segurança dos dados** – possuem sistemas de armazenamento de dados seguro e realizam procedimentos de *backups*;
- **disponibilidade** – possibilitam a disponibilização *on-line* de dados de pesquisa, de modo que esses sejam acessados, baixados e analisados;
- **curadoria digital** – configuram-se como ambientes adequados para o desempenho de processos relacionados à curadoria de dados;
- **serviços inovadores** – possibilitam a criação de novos serviços direcionados a pesquisadores, gestores de dados e financiadores de pesquisas, por meio da análise e integração de dados;

- **reuso dos dados** – amplia a prática de reuso de dados, permitindo, dessa forma, a realização de novos estudos e a redução de custos relacionados à produção de dados;
- **redes de repositórios** – possibilidade da formação de redes de repositórios de dados, por meios de protocolos de interoperabilidade, tais como o OAI-PMH e;
- **indicador de qualidade e produtividade da instituição** – as coleções de dados disponibilizadas em repositórios dão indicativos da qualidade e importância das atividades de pesquisa realizadas em instituições.

Após a exposição de aspectos relacionados à função e benefícios dos repositórios de dados de pesquisa, a seção seguinte discutirá sobre diferentes tipos de repositórios.

2.6.2 Tipos de repositórios de dados de dados de pesquisa

Considerando a diversidade de repositórios de dados de pesquisa existentes, estes podem ser classificados, conforme a literatura científica e técnica, como sendo: institucionais; disciplinares; multidisciplinares; repositórios de projetos específicos; e próprios.

2.6.2.1 Repositórios institucionais de dados de pesquisa

De acordo com Pampel *et al.* (2013), os repositórios institucionais de dados de pesquisa são geridos por uma instituição como uma universidade ou um instituto de pesquisa. No contexto universitário, abrange escopo multidisciplinar.

Conforme explica Silva (2019), os repositórios institucionais de dados de pesquisa tornaram-se alvos da atenção da comunidade científica, principalmente no início dos anos 2000, com a aparição de *softwares* que possibilitaram sua implementação, tais como Fedora e *Dspace*. São construídos com a finalidade de coletar, gerenciar e preservar a produção científica de instituições acadêmicas ou de entidades de pesquisas.

Tais repositórios contribuem para o acesso aberto na via verde, na medida em que viabilizam um meio para que pesquisadores realizem o auto-arquivamento de suas publicações. Possibilitam o armazenamento de dados, sua descrição e atribuição de identificadores necessários à sua recuperação e citação. A título de exemplo dessa categoria de repositório, pode-se destacar o repositório de dados da Universidade de Bristol, chamado *data.bris Research Data Repository* (SILVA, 2019).

Considerando o intenso ritmo da produção científica interdisciplinar ou temática no âmbito da Universidade de Bristol, o *data.bris Research Data Repository*³² viabiliza um caminho para que pesquisadores da instituição compartilhem dados abertos de pesquisa com o mundo. Esse é um repositório de dados multidisciplinar que reúne a produção intelectual da universidade. A cada conjunto de dados depositado é atribuído um único *Digital Object Identifier* (DOI).

2.6.2.2 Repositórios disciplinares de dados de pesquisa

De acordo com Silva (2019, p. 95), repositórios disciplinares de dados de pesquisa – também chamados de temáticos – abrangem os dados científicos de um campo disciplinar específico, e, geralmente, dispõem de ferramentas analíticas e de descobertas para facilitar a recuperação de dados. Nessa direção, várias disciplinas lançam mão de repositórios singularmente arquitetados para comportar os tipos de dados característicos de seus domínios, como se pode observar:

São diversas as disciplinas que dispõem de repositórios desenhados especificamente para os tipos de dados de seu domínio. Alguns exemplos são: o Protein Data Bank da Research Collaboratory for Structural Bioinformatics (RCSB), para plataformas 3D das proteínas, ácidos nucleicos e conjuntos complexos; o GenBank, para as sequências de ADN; o EMDataBank, com mapas 3D de microscopia eletrônica de densidade, modelos atômicos e metadados associados; o eCrystals, para dados cristalográficos de raios X, e a National Oceanographic Data Center (NODC), para dados oceanográficos (SILVA, 2019, p. 95).

32 Disponível em: <https://data.bris.ac.uk/data/about>

2.6.2.3 Repositórios multidisciplinares de dados de pesquisa

Conforme o entendimento de Pampel *et al.* (2013), repositórios multidisciplinares de dados de pesquisa atendem a demandas referentes a diversas áreas do conhecimento, não somente a um campo disciplinar.

O *Figshare*³³ é um exemplo de repositório de dados multidisciplinar. Foi projetado para a gestão e disseminação de dados de pesquisa gerados em ambiente acadêmico. Tecnicamente apropriado para manipular diferentes formatos de arquivos, esse repositório hospeda e preserva dados ao longo do tempo. A respeito do *Figshare*, Silva (2019) comenta:

Figshare é uma plataforma criada por *Digital Science* que permite compartilhar e mostrar os resultados de pesquisas multidisciplinares e que está dirigida a pesquisadores, cientistas, projetos e instituições. Atualmente está associada com *F1000 Research* (um prestigioso repositório de artigos científicos), colabora com *PLOS* (a maior revista científica de acesso aberto do mundo) e também com *Plum Analytics* (um serviço que quantifica o impacto dos trabalhos de pesquisa publicados). Todo o material publicado em *Figshare* é identificado com um DOI para facilitar sua localização e sua citação. Na plataforma podemos localizar: apresentações, vídeos, pôsteres, imagens, dados, artigos etc., e a preservação dos dados funciona com tecnologia *CLOCKSS*, uma organização sem fins lucrativos que promove a aliança entre os editores do mundo acadêmico e as bibliotecas acadêmicas para arquivar de um modo sustentável todo o conteúdo web produzido no âmbito científico (SILVA, 2019, *grifo do autor*).

2.6.2.4 Repositórios de dados de pesquisa de projetos específicos

Para Pampel *et al.* (2013), os repositórios de dados de pesquisa de projetos específicos concentram-se no armazenamento e disponibilização de dados decorrentes de projetos de pesquisa específicos. Como ilustração dessa categoria de repositório, pode-se destacar o *Digital Pantheon*³⁴, que disponibiliza dados de pesquisa gerados no projeto intitulado *Bern Digital Pantheon*. Esse, por seu turno, logrou êxito na disponibilização de imagens do Panteão digitalizadas em 3D por meio de *scanner a laser*.

33 Disponível em: <https://help.figshare.com/article/what-is-figshare>

34 Disponível em: <http://repository.edition-topoi.org/collection/BDPP/overview>

2.6.2.5 Repositórios de dados de pesquisa próprios

Silva (2019) explica que, em determinadas situações, pesquisadores armazenam dados provenientes de suas pesquisas em algum servidor pessoal ou em nuvem. Nessa direção, há opções tecnológicas gratuitas como *Dropbox*, e outras executadas por meio de pagamento, como a *Amazon Cloud Drive*. Em razão de a manutenção desses repositórios depender dos próprios pesquisadores para os adaptar conforme as demandas de seus projetos e para providenciar ações de *backup*, não há garantias de que os conjuntos de dados sejam devidamente organizados e preservados a longo prazo.

A próxima seção dedicar-se-á à apresentação de repositórios de dados de pesquisa que se destacam no contexto brasileiro, sendo esses frutos de iniciativas exclusivamente nacionais ou resultados de colaboração científica com outros países.

2.6.3 Repositórios de dados de pesquisa no contexto brasileiro

Para ilustrar as diferentes possibilidades temáticas, tecnológicas e de tipologias de dados que armazenam os repositórios de dados de pesquisa no contexto brasileiro, esse trabalho recorreu a informações disponibilizadas no *site* da iniciativa re3data.

Financiado pela *German Research Foundation* (DFG), re3data³⁵ é um registro mundial de repositórios de dados de pesquisa em diferentes áreas do conhecimento. Reúne repositórios que armazenam permanentemente conjuntos de dados, promove a cultura de compartilhamento desses e amplia a visibilidade de repositórios. Dessa forma, são bastante úteis a pesquisadores, editores científicos e instituições acadêmicas que buscam por repositórios apropriados para o depósito dos dados que produzem.

Considerando o exposto, uma busca no re3data foi realizada em 16 de fevereiro de 2021, na qual foram recuperados onze repositórios de dados de pesquisa de origem brasileira ou decorrentes de consórcios internacionais dos quais o Brasil

35 Disponível em: <https://www.re3data.org/faq>

participa como membro. Tais repositórios podem ser brevemente descritos da seguinte maneira:

- **WorldClim - Global Climate Data** – é um banco de dados sobre camadas climáticas globais de alta resolução espacial, cujos dados podem ser usados para mapeamento e modelagem espacial, por meio de programas de computador.
- **IBICT Cariniana Dataverse Network (Instituto Brasileiro de Informação em Ciência e Tecnologia Cariniana Dataverse Network)** – criado pelo Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT), é um repositório de dados de pesquisa multidisciplinar que oferece meios gratuitos para o depósito, disponibilização e recuperação de conjuntos de dados armazenados pelas instituições participantes da rede Cariniana. Segundo informações disponíveis no *site* do IBICT³⁶, a Rede Cariniana surgiu com a finalidade de prestar um serviço de preservação digital de documentos e de acesso continuado à produção científica armazenada digitalmente no Brasil. Dentre os serviços oferecidos pela rede, destacam-se a preservação digital de teses e dissertações, e repositórios de dados de pesquisa.
- **International Ocean Discovery Program (IODP)** – é um programa de pesquisa marinha internacional que estuda a história e a dinâmica da Terra, por meio de plataformas de pesquisa oceânicas. Essas possibilitam a recuperação de dados relacionados a rochas localizadas no fundo do mar, e, por conseguinte, o monitoramento de ambientes submarinos.
- **GLOBE (Global Collaboration Engine)** – é um espaço colaborativo *online* que viabiliza aos pesquisadores o compartilhamento, acesso a dados e realização de estudos comparativos relacionados às mudanças ocorridas em sistemas terrestres ao longo do tempo.
- **FAPESP COVID-19 Data Sharing/BR** – repositório que engloba dados de pacientes testados para COVID-19 (com resultados positivos ou negativos para a enfermidade) em instituições de saúde brasileiras. A partir dessa iniciativa,

são disponibilizados três tipos de dados: dados demográficos; provenientes de exames clínicos; laboratoriais. Além disso, são fornecidos dados atualizados periodicamente sobre hospitalização de pacientes.

- **PPBio Data Repository (Repositório de Dados de Levantamentos Biológicos)** – foi criado em 2004, com a finalidade de estimular a produção científica sobre biodiversidade no Brasil, integrar práticas de pesquisa e disseminar resultados de estudos, principalmente, na área de gestão e educação ambiental. Ademais, o PPBio colabora com a rede *DataONE*, a partir do compartilhamento dos dados que produz.
- **Open Research Data @PUC-Rio (Coleção Digital – Dados de Pesquisa @PUC-Rio)** – é uma plataforma agregadora criada para proporcionar acesso facilitado a dados de pesquisa e a outros conteúdos digitais do Repositório Maxwell. Nesse agregador, todos os conjuntos de dados são licenciados sob uma Licença CC.
- **CEDAP Research Data Repository (Centro de Documentação e Acervo Digital da Pesquisa – Dados de Pesquisa)** – criado em 2017, pela Universidade Federal do Rio Grande do Sul (UFRGS), esse repositório tem o objetivo de agrupar dados de pesquisa coletados em diversos campos disciplinares. Suas atividades abrangem o desenvolvimento de políticas, planejamento, gestão, descrição, armazenamento, compartilhamento e reuso de dados de pesquisa.
- **Base de Dados Científicos da Universidade Federal do Paraná (Scientific Database of the Federal University of Paraná)** – cumpre o propósito de agrupar dados de pesquisa produzidos pela comunidade da UFPR, o que inclui teses e dissertações, artigos científicos, entre outros recursos de informação. Acompanhando a tendência internacional da criação de produtos e serviços voltados à gestão, organização, armazenamento, compartilhamento e reuso de pesquisas, esse repositório contribui principalmente para o aumento da visibilidade da produção científica da universidade.
- **Exploration and Production Data Bank (BDEP)** – foi criado em 2000, com o objetivo de armazenar e disponibilizar informações geofísicas, geológicas e geoquímicas. É uma base de dados que auxilia pesquisadores do campo de

bacias sedimentares na identificação de áreas nas quais há maior probabilidade de se encontrar petróleo e gás natural.

- **Repositório de dados de pesquisa do Instituto Federal de Goiânia – Campus Urutaí (Research Data Repository of the Federal Institute of Goiania)** – foi construído com a finalidade de disponibilizar dados de pesquisa, de modo que esses tenham sua autoria devidamente atribuída. O reuso de dados está previsto para ser incorporado às atividades didático-pedagógicas realizada no campus, bem como para a condução de novos estudos.

À vista do exposto, igualmente são importantes discussões acerca de metadados para dados de pesquisa. Isso porque tais elementos facilitam a localização, recuperação e reuso de dados. Considerando os aspectos mencionados, a próxima seção terá como tópico central a apresentação de estudos sobre metadados utilizados no contexto de dados de pesquisa.

2.7 Metadados para dados de pesquisa

De acordo com informações coletadas no *site* da *OpenAIRE*³⁷, metadados são dados que provêm informações sobre dados, tornando-os localizáveis, recuperáveis e reutilizáveis. Podem incluir informações como localização geográfica, unidades de medida, abreviações utilizadas em conjuntos de dados, detalhes sobre instrumentos e protocolos adotados na pesquisa, entre outras. Também podem apresentar diferentes formas, o que abrange texto livre, conteúdo padronizado e legível por máquina. Assim sendo, é altamente recomendável a utilização de um formato de metadados padronizado conforme o campo disciplinar abordado.

Segundo a Universidade de Oulu³⁸, na Finlândia, metadados são entendidos como dados estruturados e legíveis por computadores, que tratam a respeito de dados. Nessa perspectiva, “qualquer arquivo de dados em qualquer formato deve ter campos de metadados, pois sem metadados e os conjuntos de dados de

37 Disponível em: <https://www.openaire.eu/what-is-metadata>

38 Disponível em: https://libguides oulu.fi/Researchdata/Data_documentation

documentação, não faz sentido”. Portanto, o principal objetivo dos metadados é auxiliar a localização de dados de pesquisa.

Para a Universidade de *New England* (UNE)³⁹, os metadados referem-se a informações sobre coleções de dados, o que abrange aspectos como seu conteúdo, qualidade, metodologias utilizadas, permissões de uso e requisitos de gestão. Na Universidade, os metadados para dados de pesquisa são definidos por bibliotecários de dados de pesquisa, juntamente com o pesquisador. Após o preenchimento de todos os campos mínimos de metadados e a conclusão do registro de metadados, esse poderá ser acessado por meio do repositório institucional da UNE. A título de ilustração, o Quadro 7 mostra os campos mínimos de metadados adotados no registro da UNE:

Quadro 7 – Campos mínimos de metadados para o registro

ABA DE REGISTRO	Campo de registro
GERAL	Título Tipo
COBERTURA	Cobertura de data Localização geoespacial
DESCRIÇÃO	Descrição do conjunto de dados - resumo, metodologia e fontes Publicação relacionada Websites relacionados
PESSOAS	Criadores Contato primário
ASSUNTO	Códigos de campo de pesquisa Códigos de objetivo socioeconômico Palavras-chave para o assunto da pesquisa
DIREITOS	Direitos de acesso Licença de uso
GESTÃO	Identificador para o conjunto de dados URL para o armazenamento de dados Período de retenção Quantidade de arquivos Contato do custodiante de dados para acesso Números de concessão associados
SUBMISSÃO	Data de embargo

Fonte: Traduzido e adaptado de UNE

39 Disponível em: <https://www.une.edu.au/research/digital-research-support/research-data-management/metadata-for-research-data>

Considerando as múltiplas realidades nas quais estão inseridos os metadados para dados de pesquisa, a próxima seção abordará as categorias conceituais em que esses podem ser classificados.

2.7.1 Categorias de metadados

Conforme o entendimento de Sayão (2010), a classificação de metadados em categorias não é um consenso entre especialistas. Entretanto, uma parcela bastante representativa dos autores afirma que os metadados podem ser agrupados nas seguintes categorias:

- **metadados descritivos** – constituindo a categoria mais conhecida, esses metadados descrevem um recurso com a finalidade de facilitar sua localização e identificação. Podem abranger características como título, autor e palavras-chave;
- **metadados estruturais** – são informações que registram como devem ser organizados os recursos constituídos por vários elementos. A título de exemplo, pode-se destacar a ordenação e vinculação lógica de páginas de um livro que foram digitalizadas separadamente;
- **metadados administrativos** – são informações que apoiam os processos de gestão de recursos informacionais, como aquelas relacionadas à data, ao modo e ao porquê de sua criação. Essa categoria também engloba os chamados metadados técnicos, os quais apresentam as especificidades técnicas do recurso, e os metadados que auxiliam a gestão relativa a direitos.

Semelhantemente, Sanchez, Silva e Vechiato (2018) explicam que os metadados descritivos permitem que recursos sejam recuperados mediante utilização de mecanismos de busca. Já os metadados administrativos, são aqueles que indicam aspectos relativos às origens de um recurso, à cobertura de seus direitos autorais e intelectuais, entre outros. Por seu turno, os metadados estruturais viabilizam a correlação entre recursos.

Dando prosseguimento à discussão relacionada aos metadados para dados de pesquisa, a seção seguinte discorrerá sobre padrões de metadados.

2.7.2 Padrões de metadados

Segundo o *Digital Curation Centre* (DCC), muitas disciplinas apoiam a elaboração de especificações ou padrões de metadados, os quais a comunidade científica considera necessários para o reuso de dados. Com base em diretório do DCC⁴⁰, serão brevemente descritos padrões de metadados disciplinares na área de Ciências Sociais e Humanidades, Ciência Física, Ciências da Terra e Biologia, que ilustram a diversidade desses padrões. Com relação a disciplinas que ainda não definiram um padrão de metadados ou lidam com dados oriundos de várias áreas do conhecimento, serão abordados padrões de dados gerais de pesquisa.

Ciências Sociais e Humanidades

- **DDI - Data Documentation Initiative**⁴¹ – padrão internacional amplamente conhecido na descrição de dados nas áreas das ciências sociais, comportamentais e econômicas. Há duas versões sendo mantidas: o *DDI Codebook* (ou DDI versão 2) (destinado à documentação de dados menos complexos) e o *DDI Lifecycle* (ou DDI versão 3) (destinado à documentação de conjuntos de dados complexos em cada estágio de seu ciclo de vida, desde a etapa em que foram gerados, até a publicação e reuso);
- **MIDAS-Heritage**⁴² – padrão utilizado para o registro de informações sobre edificações, sítios arqueológicos, jardins, campos de batalha e artefatos, como forma de documentar aspectos relativos ao patrimônio cultural britânico;
- **QuDEx - Qualitative Data Exchange Format**⁴³ – consiste em um formato neutro de *software* para dados qualitativos que mantém as anotações e os relacionamentos entre dados e outros objetos;

40 Disponível em: <https://www.dcc.ac.uk/guidance/standards/metadata>

41 Disponível em: <https://www.dcc.ac.uk/resources/metadata-standards/ddi-data-documentation-initiative>

42 Disponível em: <https://www.dcc.ac.uk/resources/metadata-standards/midas-heritage>

43 Disponível em: <https://www.dcc.ac.uk/resources/metadata-standards/qudex-qualitative-data-exchange-format>

- **SDMX - Statistical Data and Metadata Exchange**⁴⁴ – orientações formais, diretrizes técnicas e estatísticas utilizadas para o intercâmbio e compartilhamento de dados e metadados estatísticos.

Ciência Física

- **AVM - Astronomy Visualization Metadata**⁴⁵ – suporta a busca em coleções de imagens astronômicas obtidas a partir de observações telescópicas. O esquema possibilita que metadados sejam incorporados em formatos de imagem, tais como JPEG, TIFF e PNG;
- **CIF - Crystallographic Information Framework**⁴⁶ – estrutura de arquivo padrão utilizada para o arquivamento e compartilhamento de informações cristalográficas;
- **International Virtual Observatory Alliance Technical Specifications (IVOA)**⁴⁷ – recomendações técnicas estabelecidas pela IVOA viabilizam a interoperabilidade e combinação de arquivos astronômicos gerados em todo o mundo. Tais arquivos dispõem de diferentes modelos que funcionam como esquemas de metadados para conjuntos de dados específicos, como dados de fotometria, espectrais e observacionais;
- **NeXus**⁴⁸ - padrão internacional para armazenamento e intercâmbio de dados obtidos em experimentos de nêutrons, raios-x e *múons*. A estrutura dos arquivos NeXus é flexível e permite o armazenamento de conjuntos de dados simples e complexos, juntamente com seus metadados associados. Além disso, NeXus possui regras específicas de domínios disciplinares para organizar dados em arquivos HDF5.

44 Disponível em: <https://www.dcc.ac.uk/resources/metadata-standards/sdmx-statistical-data-and-metadata-exchange>

45 Disponível em: <https://www.dcc.ac.uk/resources/metadata-standards/avm-astronomy-visualization-metadata>

46 Disponível em: <https://www.dcc.ac.uk/resources/metadata-standards/cif-crystallographic-information-framework>

47 Disponível em: <https://www.dcc.ac.uk/resources/metadata-standards/international-virtual-observatory-alliance-technical-specifications>

48 Disponível em: <https://www.dcc.ac.uk/resources/metadata-standards/nexus>

Ciências da terra

- **AgMES - Agricultural Metadata Element Set**⁴⁹ - padrão semântico elaborado pela *Food and Agriculture Organization (FAO)*, o AgMES possibilita a descrição, a descoberta e o intercâmbio de dados em áreas relacionadas à produção de alimentos, nutrição humana e desenvolvimento rural;
- **CF (Climate and Forecast) Metadata Conventions**⁵⁰ - foi criado como um padrão para dados disponíveis no formato *netCDF*, com dados relacionados à previsão climática. Todavia, igualmente pode ser aplicado ao contexto de dados observacionais e disponíveis em outros formatos.
- **DIF - Directory Interchange Format**⁵¹ - um dos primeiros padrões de metadados para a descrição de dados de pesquisa no campo das Ciências da Terra. Abrange elementos que focam em instrumentos de captura de dados e nas próprias características dos dados.
- **ISO 19115**⁵² - padrão internacionalmente utilizado para descrever informações geográficas, abrangendo características como a extensão, e a referência espacial e temporal de dados geográficos digitais.

Biologia

- **ABCD - Access to Biological Collection Data**⁵³ - suportando dados de diversas fontes informacionais, esse padrão viabiliza o acesso e o intercâmbio de dados sobre espécimes (também chamados de dados primários de biodiversidade);
- **Darwin Core**⁵⁴ - conjunto de padrões e um glossário de termos utilizados para facilitar o compartilhamento de informações relativas à diversidade biológica;

49 Disponível em: <https://www.dcc.ac.uk/resources/metadata-standards/agmes-agricultural-metadata-element-set>

50 Disponível em: <https://www.dcc.ac.uk/resources/metadata-standards/cf-climate-and-forecast-metadata-conventions>

51 Disponível em: <https://www.dcc.ac.uk/resources/metadata-standards/dif-directory-interchange-format>

52 Disponível em: <https://www.dcc.ac.uk/resources/metadata-standards/iso-19115>

53 Disponível em: <https://www.dcc.ac.uk/resources/metadata-standards/abcd-access-biological-collection-data>

54 Disponível em: <https://www.dcc.ac.uk/resources/metadata-standards/darwin-core>

- **EML - Ecological Metadata Language**⁵⁵ – baseado no trabalho realizado pela *Ecological Society of America*, esse padrão de metadados foi criado especialmente para o campo disciplinar da Ecologia;
- **Genome Metadata**⁵⁶ – padrão constituído por 61 campos de metadados diferentes, denominados atributos. Esses, são distribuídos em sete categorias gerais (informações do organismo, informações do isolamento, informações do hospedeiro, informações do sequenciamento, informações do fenótipo, informações do projeto e outros).

Dados Gerais de pesquisa

- **CERIF - Common European Research Information Format**⁵⁷ – padrão recomendado pela União Europeia aos seus estados membros em relação ao registro de informações sobre pesquisas. A partir da versão 1.6, acrescentou suporte específico a metadados para conjuntos de dados.
- **Data Package**⁵⁸ – é um padrão genérico para intercâmbio de dados. Define um formato com campos obrigatórios, sugeridos e opcionais;
- **DataCite Metadata Schema**⁵⁹ – quando um *Digital Object Identifier* (DOI) for criado para determinado conjunto de dados, metadados obrigatórios devem ser registrados no *DataCite Metadata Store*, principalmente por esses metadados auxiliarem na identificação específica e persistente de dados;
- **Dublin Core**⁶⁰ – recorrente em diferentes domínios, esse é um dos mais conhecidos e adotados padrão de metadados existentes.

55 Disponível em: <https://www.dcc.ac.uk/resources/metadata-standards/eml-ecological-metadata-language>

56 Disponível em: <https://www.dcc.ac.uk/resources/metadata-standards/genome-metadata>

57 Disponível em: <https://www.dcc.ac.uk/resources/metadata-standards/cerif-common-european-research-information-format>

58 Disponível em: <https://www.dcc.ac.uk/resources/metadata-standards/data-package>

59 Disponível em: <https://www.dcc.ac.uk/resources/metadata-standards/datacite-metadata-schema>

60 Disponível em: <https://www.dcc.ac.uk/resources/metadata-standards/dublin-core>

3 METODOLOGIA

De acordo com Nascimento (2012), a metodologia é uma ciência instrumental que trata da aplicação de métodos na pesquisa científica, de modo a fornecer caminhos ou alternativas na resolução de problemas os quais pesquisadores pretendem resolver. Em abordagem similar à de Nascimento (2012), Tomanick (2004) destaca que problemas de pesquisa devem ser solucionados por cientistas a partir da escolha dos métodos mais adequados ao contexto do estudo realizado. O método, por seu turno, pode ser entendido na perspectiva de Appolinário (2011), em sua obra intitulada *Dicionário de metodologia científica: um guia para a produção do conhecimento científico*, como parte de um trabalho científico na qual são descritos os procedimentos adotados em determinado estudo, tais como a seleção de participantes e a escolha de instrumentos para a coleta de dados.

À luz do entendimento desses autores sobre metodologia e método, principalmente sob o espectro da descrição dos procedimentos adotados na pesquisa científica, as seções 3.1 a 3.3 apresentarão os caminhos metodológicos tomados no decorrer deste estudo, referindo-se ao modelo conceitual da pesquisa, à caracterização da pesquisa e aos procedimentos metodológicos.

3.1 Modelo conceitual da pesquisa

Segundo Sekaran (2003), o modelo conceitual (também conhecido por quadro conceitual) oferece a base de conceitos relacionados que estrutura toda a pesquisa em nível teórico. Na medida em que existem diferentes entendimentos sobre o termo “modelo”, o autor esclarece que utiliza esse termo no sentido de uma representação de relações entre conceitos, ou seja, como um esquema conceitual que associa conceitos. Nessa perspectiva, o modelo conceitual representa as inter-relações entre as variáveis mais importantes para a dinâmica do problema investigado.

O modelo conceitual é composto por uma rede de conceitos logicamente desenvolvidos e identificados a partir de processos como a realização de entrevistas e o levantamento da literatura (SEKARAN, 2003). A importância de o pesquisador

recorrer à literatura antes de conceber o modelo conceitual da pesquisa é evidenciada nos seguintes argumentos de Sekaran (2003, p. 97):

A relação entre o levantamento da literatura e o modelo conceitual é que o primeiro fornece uma base sólida para o desenvolvimento do último. Ou seja, o levantamento da literatura identifica as variáveis que podem ser importantes, conforme determinado por resultados de pesquisas anteriores. Isso, além de outras conexões lógicas que podem ser conceituadas, forma a base para o modelo conceitual. O modelo conceitual elabora as relações entre as variáveis, explica a teoria subjacente a essas relações e descreve a natureza e a direção das relações. Assim como o levantamento da literatura prepara o terreno para um bom arcabouço teórico, este, por sua vez, fornece a base lógica para o desenvolvimento de hipóteses testáveis (SEKARAN, 2003, p. 97, *tradução nossa*).

Tendo como pano de fundo os argumentos de Sekaran (2003) apresentados sobre a relação de conceitos relevantes que norteiam teoricamente uma pesquisa, afirma-se que a teoria embutida no modelo conceitual do presente estudo pressupõe que o entendimento das diferenças entre as áreas é imprescindível na elaboração de diretrizes para políticas institucionais de gestão de dados abertos de pesquisa em universidades. Isso porque, nessas instituições, pode-se observar com nitidez como se manifesta a pluralidade das áreas do conhecimento em relação a diversos aspectos, como os modos pelos quais essas áreas produzem e comunicam dados e resultados de pesquisa.

Além disso, o estudo compreende que a gestão de dados de pesquisa se dá mediante a realização de um conjunto de práticas indissociavelmente relacionadas a etapas que constituem o ciclo de vida de dados, que, por seu turno, são etapas fortemente influenciadas pelas diferenças entre as áreas do conhecimento. Conforme essas explicações, a Figura 16 apresenta o modelo conceitual da pesquisa.

Figura 16 – Modelo conceitual da pesquisa



Fonte: elaborado pela autora

Lançando mão de modelos referidos na literatura científica e técnica (UNIVERSIDADE DE VIRGINIA; DATAONE; UK DATA ARCHIVE; USGS; ICPSR; GFBIO; JISC; DTU; RÜEGG *et al.*), observa-se que modelos do ciclo de vida de dados distinguem-se entre si quanto à nomenclatura e ordem em que são apresentadas suas etapas, aos atores envolvidos nos processos, aos contextos disciplinares os quais pertencem os dados, dentre outros aspectos. Para fins deste trabalho, a gestão dos dados abertos de pesquisa é diretamente influenciada por diferenças entre as áreas do conhecimento e abrange as seguintes etapas do ciclo de vida de dados:

- **planejar** – etapa que antecede a realização da pesquisa. Engloba o planejamento de atividades e processos que serão executados durante o ciclo de vida de dados de pesquisa, o que inclui: a definição dos objetivos do estudo; a previsão de custos; a estimativa de tempo e elaboração de cronograma; o estabelecimento de requisitos de qualidade dos dados e de padrões para a produção, compartilhamento, reuso e arquivamento de dados. Nessa etapa, o pesquisador também elabora o Plano de gestão de dados de pesquisa, que, por seu turno, deve ser atualizado de acordo com o andamento da pesquisa;
- **coletar** – engloba as técnicas de coleta de dados utilizadas pelo pesquisador. Conjuntos de dados são variáveis em relação à natureza, aos formatos e aos modos pelos quais foram produzidos e/ou obtidos, principalmente devido a influência de diferenças disciplinares;
- **processar** – refere-se ao processamento técnico, à organização e à documentação de dados. Inclui atividades como a transcrição, anonimização e tradução de dados. A documentação engloba a descrição detalhada de dados, conforme padrões e protocolos determinados, com o propósito de torná-los compreensíveis e recuperáveis no futuro. Trata-se, ainda, da utilização de metadados previamente escolhidos para caracterizar aspectos relevantes de dados, tal como o contexto em que foram gerados e analisados;
- **compartilhar** – compreende o compartilhamento de dados por meio de canais informais, como o intercâmbio de dados via *e-mail* e/ou por canais formais, como o depósito de dados em repositórios e em periódicos de dados. Nessa etapa, são estabelecidos os níveis de acesso aos dados e as licenças para a disponibilização de dados;
- **preservar** – conjunto de atividades realizadas com vistas à preservação de dados a longo prazo, mantendo-os íntegros e disponíveis no futuro. Inclui a realização de *backups*, a migração e atualização de formatos, o armazenamento seguro de dados e de metadados, e a elaboração do Plano de Preservação de dados;

- **descobrir** – adoção de estratégias que oferecem maior visibilidade a dados no contexto da pesquisa científica. Nessa etapa, dados são descobertos por meio de catálogos de serviços de dados, metadados utilizados em repositórios de dados, periódicos de dados, entre outros;
- **reutilizar** – etapa em que dados são reutilizados em pesquisas, possibilitando o desencadeamento de novos trabalhos científicos, a integração de dados a estudos realizados e a possibilidade de análises secundárias, a replicação e a apuração de resultados de pesquisas;
- **monitoramento e controle** – etapa contínua durante o ciclo de vida de dados. O monitoramento refere-se ao acompanhamento e avaliação de processos e atividades da pesquisa, com vistas à qualidade nos padrões e protocolos adotados. Já o controle, está relacionado a ações corretivas em falhas identificadas em relação aos dados no decorrer de seu ciclo;

No referido modelo, três conceitos retirados da literatura científica e técnica são fundamentais – diferenças entre as áreas do conhecimento; gestão de dados abertos; diretrizes para a gestão de dados abertos de pesquisa. À título de operacionalização conceitual, esses serão definidos nos tópicos seguintes. Portanto, para fim dessa pesquisa, entende-se por:

- **diferenças entre as áreas do conhecimento**⁶¹ – as especificidades inerentes a cada área do conhecimento, no que se refere à própria natureza da área, às suas tradições em termos de dados de pesquisa, às abordagens metodológicas adotadas e aos modos como produzem e comunicam os resultados de pesquisa. Logo, tais especificidades evidenciam as diferenças das áreas entre si;
- **gestão de dados abertos de pesquisa** – conjunto de práticas contínuas que viabilizam a preservação e reuso de dados abertos de pesquisa, ao longo de todo o seu ciclo de vida. Sob a perspectiva de acesso aberto, entende-se que

61 A definição apresentada provém de revisão de literatura, que, por seu turno, recorrentemente utiliza o termo “diferenças disciplinares” para se referir à distinção entre disciplinas decorrente do processo evolutivo da especialização das ciências. Porém, para fins de operacionalização da pesquisa, foi considerado um estágio classificatório mais genérico, que remete às diferenças entre as áreas do conhecimento. Nesse sentido, as áreas consideradas no estudo serão a Sociologia e a Química, determinadas, prioritariamente, por uma questão de recorte amostral.

dados abertos de pesquisa são dados que podem ser livremente compartilhados e reutilizados sem quaisquer barreiras, por quaisquer indivíduos, desde que reconhecida a autoria dos dados;

- **diretrizes para a gestão de dados abertos de pesquisa**⁶² – instruções formais que servem de orientação para a gestão de dados abertos de pesquisa. Tendo em vista o contexto de universidades o qual se direciona o presente estudo, trata-se do conjunto de diretivas para a gestão de dados abertos em diferentes áreas do conhecimento, de modo a considerar as etapas que constituem o ciclo de vida de dados de pesquisa.

3.2 Caracterização da pesquisa

De acordo com Creswell (2014), as abordagens metodológicas adotadas na pesquisa científica são orientadas por crenças e pressupostos filosóficos trazidos por pesquisadores, estando esses conscientes disso, ou não. Tais pressupostos resultam do contato de pesquisadores com a literatura científica e com a comunidade acadêmica e, por conseguinte, influenciam fortemente a condução de uma pesquisa. Nesse contexto, o autor detalha quatro pressupostos filosóficos comuns à pesquisa qualitativa: ontológico, epistemológico, axiológico e metodológico.

Para Creswell (2014), o pressuposto ontológico caracteriza-se, sobretudo, pelo reconhecimento de múltiplas realidades vivenciadas por pesquisadores, participantes de pesquisas e leitores. A título de exemplo, o autor reporta-se à compilação de uma fenomenologia, na qual escritores relatam as diferentes perspectivas de participantes em relação a determinado fenômeno estudado. Por outro lado, estudos orientados pelo pressuposto epistemológico partem do princípio no qual o conhecimento sobre algo é obtido por meio das experiências subjetivas das pessoas. Por essa razão, pesquisadores geralmente coletam dados no ambiente onde os participantes da pesquisa vivem, de modo a tentar chegar o mais próximo das visões de mundo que esses possuem. Citando caso análogo, destaca-se a realização de etnografias, nas

62 Definição adotada com base no significado do verbete “diretrizes”, apontado no Léxico: dicionário de português *online*. Disponível em: <https://www.lexico.pt/diretrizes/>.

quais pesquisadores convivem no contexto de participantes da pesquisa durante determinado período (CRESWELL, 2014).

O pressuposto axiológico, por seu turno, enfatiza a incorporação de valores pessoais do pesquisador em seus estudos. Acrescenta-se a essa questão, a importância de pesquisadores se posicionarem criticamente sobre determinada situação, tendo em vista que os dados coletados na pesquisa de campo estão imbuídos de sua interpretação. Como exemplo, destaca-se a realização de uma biografia interpretativa, na qual é evidente a atribuição de valores e significados do pesquisador em texto biográfico (CRESWELL, 2014).

Por fim, Creswell (2014, p. 34) define os pressupostos metodológicos como “indutivos, emergentes e moldados pela experiência do pesquisador na coleta e análise dos dados”, em razão de três aspectos principais. O primeiro é que o pesquisador segue a lógica indutiva na condução da pesquisa. O segundo é que, algumas vezes, as questões de pesquisa inicialmente propostas modificam-se para aumentar a compreensão do pesquisador sobre o problema investigado. Como efeito disso, a técnica de coleta de dados modifica-se para acompanhar as novas questões elaboradas no estudo. O terceiro refere-se à análise de dados na qual o pesquisador busca construir um conhecimento minucioso a respeito do tópico de pesquisa investigado.

O Quadro 8 apresenta uma síntese dos pressupostos filosóficos propostos por Creswell (2014), de maneira a considerar as principais questões, características e implicações para a prática de cada um desses pressupostos.

Quadro 8 - Pressupostos filosóficos com implicações para a prática

PRESSUPOSTOS	QUESTÕES	CARACTERÍSTICAS	IMPLICAÇÕES PARA A PRÁTICA (EXEMPLOS)
Ontológico	Qual é a natureza da realidade?	A realidade é múltipla quando vista por meio de múltiplas perspectivas	O pesquisador relata diferentes perspectivas à medida que os temas se desenvolvem nos achados
Epistemológico	O que conta como conhecimento? Como as afirmações de conhecimento são justificadas? Qual é a relação entre pesquisador e quem está sendo pesquisado?	Evidências subjetivas dos participantes; o pesquisador tenta reduzir a distância entre ele e quem está sendo pesquisado	O pesquisador se baseia em citações como evidências do participante; colabora, passa um tempo no campo com os participantes e se torna um "incluído"
Axiológico	Qual é o papel dos valores?	O pesquisador reconhece que a pesquisa é carregada de valores e que os vieses estão presentes	O pesquisador discute abertamente os valores que moldam a narrativa e inclui a sua interpretação em conjunto com as interpretações dos participantes
Metodológico	Qual é o processo da pesquisa? Qual é a linguagem da pesquisa?	O pesquisador usa a lógica indutiva, estuda o tópico dentro do seu contexto e usa um projeto emergente	O pesquisador trabalha com particularidades (detalhes) antes das generalizações, descreve em detalhes o contexto do estudo e continuamente revisa questões das experiências no campo

Fonte: adaptado de Creswell (2014, p.33)

Creswell (2014) argumenta que os pressupostos filosóficos estão relacionados com estruturas interpretativas que direcionam o sentido da pesquisa. É interessante notar que o autor ilustra a variedade dessas estruturas, citando algumas delas como o positivismo, o pós-positivismo, o construtivismo, a teoria crítica, os modelos de pesquisa culturais, a teoria *queer*, a perspectiva transformadora e as abordagens das deficiências.

A título de ilustração, Creswell (2014) aborda a teoria fundamentada para demonstrar a aplicabilidade de duas estruturas interpretativas na pesquisa científica. De acordo com o autor, se um trabalho utiliza o método da teoria fundamentada almejando dar ênfase em uma abordagem objetiva e no modelo teórico resultante, minimizando a perspectiva do pesquisador, convém que seja utilizada uma estrutura interpretativa pós-positivista. Por outro ângulo, se um estudo narrativo investiga a luta pela identidade e pelo respeito aos valores de algum grupo marginalizado de

aprendizes com deficiência, exigindo mudanças a esse respeito, convém a aplicação da estrutura interpretativa das deficiências.

Em razão do reconhecimento da contribuição advinda de diferentes perspectivas dos participantes em relação ao fenômeno estudado, esta pesquisa adotou o pressuposto filosófico ontológico, visando atingir o objetivo de propor diretrizes para políticas institucionais de gestão de dados abertos de pesquisa em universidades. Semelhantemente, a pesquisa foi conduzida à luz da estrutura interpretativa construtivista, uma vez que enfatiza a visão de mundo de participantes sobre a realidade estudada, constrói uma teoria a partir dos dados coletados em campo e admite a influência do próprio pesquisador em relação ao contexto estudado.

No que se refere ao construtivismo social, Creswell (2014) explica que pesquisas estruturadas sobre esse fundamento se caracterizam, sobretudo, por três aspectos. O primeiro é que essas pesquisas dão ênfase no modo pelo qual participantes do estudo atribuem significados acerca do mundo. É interessante notar que tais significados são múltiplos e decorrentes de interação histórica, social e cultural. O segundo é que pesquisadores elaboram indutivamente uma teoria, ou desenvolvem um padrão de significados sobre fenômenos estudados, ao invés de iniciarem a pesquisa com base em uma teoria existente. O terceiro é a influência da percepção do próprio pesquisador e de suas vivências em relação ao contexto estudado, conforme esclarece o autor:

Os pesquisadores reconhecem que o seu próprio background molda a sua interpretação e eles “se posicionam” na pesquisa para reconhecer como a sua interpretação flui a partir das próprias experiências pessoais, culturais e históricas. Dessa forma, os pesquisadores fazem uma interpretação do que encontram, interpretação essa moldada pelas suas experiências e por sua base de conhecimentos anteriores. A intenção do pesquisador, então, é compreender (ou interpretar) os significados que os outros têm sobre o mundo. É por isso que a pesquisa qualitativa frequentemente é chamada de pesquisa “interpretativa” (CRESWELL, 2014, p. 36).

Ampliando as classificações dos tipos de pesquisa elaboradas por estudiosos na área da Metodologia da pesquisa, como Selltitz *et al.* (1980) e Babbie (2001), e com base na classificação proposta por Danhke (1989), Hernández Sampieri, Fernández

Collado e Baptista Lucio (2006) categorizam os tipos de estudos em exploratórios, descritivos, correlacionais e explicativos.

De acordo com os autores, estudos exploratórios são comumente realizados quando seu objetivo for examinar um tema ou problema de pesquisa pouco explorado, ou vagamente abordado na literatura científica. Outrossim, são úteis para familiarizar pesquisadores em relação a fenômenos relativamente desconhecidos, para sugerir postulados sobre determinada realidade, ou para a análise de viabilidade de pesquisas em relação a um contexto muito particular, a partir da obtenção de informações preliminares (HERNÁNDEZ SAMPIERI; FERNÁNDEZ COLLADO; BAPTISTA LUCIO, 2006).

Estudos descritivos referem-se a pesquisas que descrevem como se manifesta certo fenômeno, apontando-lhe suas propriedades e características, mas não estabelecendo relações entre variáveis analisadas. Por outro lado, estudos correlacionais, geralmente, relacionam duas ou mais variáveis em determinado contexto (HERNÁNDEZ SAMPIERI; FERNÁNDEZ COLLADO; BAPTISTA LUCIO, 2006).

Por fim, Hernández Sampieri, Fernández Collado e Baptista Lucio (2006, p. 107, *grifo do autor*) argumentam que estudos explicativos vão além da descrição de “conceitos, fenômenos ou do estabelecimento de relações entre conceitos, estão destinados a responder as causas dos acontecimentos, fatos, fenômenos físicos ou sociais”.

Semelhantemente, Collis e Hussey (2013) classificam a pesquisa científica conforme o seu propósito como sendo: exploratória, descritiva, analítica ou preditiva. A primeira é conduzida quando há pouco ou nenhum referencial teórico e estudos anteriores sobre o problema, ou questão de pesquisa investigada. A segunda refere-se à pesquisa conduzida para descrever os fenômenos tal como se apresentam. Em geral, é adotada para identificar e obter informações sobre características relativas a um problema investigado. A terceira, por seu turno, reportar-se ao estudo em que o pesquisador vai além de apenas descrever determinadas características, para analisar o porquê ou como ocorre o fenômeno investigado, descobrindo e mensurando relações existentes. Já a pesquisa preditiva, ocupa-se em generalizar aspectos analisados e prever a probabilidade de ocorrência de fenômenos com base em relações hipotéticas.

Fundamentado no entendimento de Collis e Hussey (2013), **o presente estudo é de natureza analítica em razão de seu objetivo** – propor diretrizes para políticas institucionais de gestão de dados abertos de pesquisa em universidades – **pressupor uma análise aprofundada de fenômenos já investigados na literatura científica, e não somente a descrição pura e a identificação de características desses fenômenos.** Além disso, o objetivo referido não abrange em seu teor a finalidade de prever a probabilidade de ocorrência de tópicos pesquisados.

De acordo com Creswell (2010), as pesquisas são orientadas por abordagens metodológicas denominadas estratégias de investigação. Tais estratégias são definidas como tipos de projetos ou modelos de métodos que direcionam os procedimentos adotados em determinado estudo. Sob essa perspectiva, as estratégias de investigação são quantitativas, qualitativas e de métodos mistos.

Hernández Sampieri, Fernández Collado e Baptista Lucio (2006) diferenciam o enfoque quantitativo do qualitativo na condução de pesquisas. Como pode ser observado em suas falas, os autores ressaltam que não há predominância de valor entre os enfoques, na medida em que ambos contribuem para o avanço do conhecimento científico:

O enfoque quantitativo utiliza a coleta e a análise de dados para responder às questões de pesquisa e testar as hipóteses estabelecidas previamente, e confia na medição numérica, na contagem e frequentemente no uso de estatística para estabelecer com exatidão os padrões de comportamento de uma população [...] [O enfoque qualitativo] está baseado em métodos de coleta de dados sem medição numérica, como as descrições e as observações. Regularmente, questões e hipóteses surgem como parte do processo de pesquisa, que é flexível e se move entre os eventos e sua interpretação, entre as respostas e o desenvolvimento da teoria. Seu propósito consiste em “reconstruir” a realidade, tal como é observada pelos atores de um sistema social predefinido. Muitas vezes é chamado de “holístico”, porque considera o “todo”, sem reduzi-lo ao estudo de suas partes. Do nosso ponto de vista, ambos os enfoques são bastante valiosos e já realizaram contribuições notáveis ao avanço do conhecimento. Nenhum é intrinsecamente melhor que o outro, são apenas diferentes em relação ao estudo de um fenômeno. Pensamos que a controvérsia entre as duas visões tem sido desnecessária e não está isenta de dogmatismo (HERNÁNDEZ SAMPIERI; FERNÁNDEZ COLLADO; BAPTISTA LUCIO, 2006).

Semelhantemente, Creswell (2010) afirma que a estratégia de investigação de métodos mistos combina abordagens quantitativas e qualitativas, de modo a usufruir pontos fortes de ambas as abordagens em um único estudo. A partir dessa

combinação, pode-se obter mais *insights* para compreender problemas de pesquisa complexos, e que muitas vezes são estudados por equipes multidisciplinares, com interesses e procedimentos metodológicos distintos.

Em razão do pressuposto filosófico ontológico e estrutura interpretativa construtivista que caracterizam este estudo, a estratégia de investigação qualitativa foi adotada. Por conseguinte, foi possível realizar uma análise vertical sobre o tema investigado, aprofundando-se em aspectos apontados por pesquisadores sobre a gestão de dados abertos de pesquisa em diferentes áreas.

Segundo Flick (2009), o termo “pesquisa qualitativa” foi utilizado durante muitos anos como uma alternativa crítica à pesquisa quantitativa, e sua ascensão remete às décadas de 1960 e 1970. Porém, no decorrer do tempo, o termo assimilou novos contextos de significância, a ponto de não ser mais definido em contraste à pesquisa quantitativa, mas sim, por suas próprias características. Assim, Flick (2009, p. 16) afirma que a pesquisa qualitativa “parte da noção da construção social das realidades em estudo, está interessada nas perspectivas dos participantes, em suas práticas do dia a dia e em seu conhecimento cotidiano relativo à questão em estudo”.

De forma semelhante a Flick (2009), Strauss e Corbin (2008) explicam que o termo “pesquisa qualitativa” refere-se a estudos sobre a vida de pessoas, especialmente em relação a suas experiências e comportamentos, a movimentos sociais e a fenômenos culturais. Do mesmo modo, apesar de algumas pesquisas lançarem mão de dados quantificáveis, o foco da análise é interpretativo.

Strauss e Corbin (2008) ainda destacam três componentes principais da pesquisa qualitativa. O primeiro diz respeito à diversidade de dados, que, por sua vez, são produzidos ou obtidos de várias fontes, como entrevistas, documentos e filmes. O segundo refere-se aos procedimentos de interpretação e organização de dados, que geralmente consistem no processo de codificação e análise dos mesmos, além elaboração de memorandos e diagramação. O terceiro é a redação de relatórios escritos, que podem ser publicados como artigos em periódicos científicos, em conferências, em livros, dentre outros canais.

Incorporando elementos presentes na definição do termo “pesquisa qualitativa” apresentada por Denzin e Lincoln (2011), Creswell (2014) define a pesquisa qualitativa

a partir de seus pressupostos e estruturas interpretativas, dando ênfase nos significados os quais indivíduos atribuem a determinado problema social ou humano, como pode ser ilustrado em suas argumentações:

A pesquisa qualitativa começa com pressupostos e o uso de estruturas interpretativas/teóricas que informam o estudo dos problemas da pesquisa, abordando os significados que os indivíduos ou grupos atribuem a um problema social ou humano. Para estudar esse problema, os pesquisadores qualitativos usam uma abordagem qualitativa da investigação, a coleta de dados em um contexto natural sensível às pessoas e aos lugares em estudo e a análise dos dados que é tanto indutiva quanto dedutiva e estabelece padrões ou temas. O relatório final ou a apresentação incluem as vozes dos participantes, a reflexão do pesquisador, uma descrição complexa e interpretação do problema e a sua contribuição para a literatura ou um chamado à mudança (CRESWELL, 2014, p. 49).

É interessante notar que Creswell (2014) ainda apresenta um conjunto de características comuns da pesquisa qualitativa, que, em síntese, são descritas sem qualquer ordem específica de importância, a saber:

- **habitat natural** – pesquisadores geralmente coletam dados no ambiente onde os participantes da pesquisa vivenciam o fenômeno ou problema investigado, e não em uma situação artificial criada em laboratório. Pesquisadores obtêm dados por meio de interações pessoais com os participantes do estudo, ouvindo-os e vendo como se comportam em determinado contexto;
- **o pesquisador como um instrumento-chave** – a influência do pesquisador, principalmente no tocante à coleta de dados, é inerente à realização da pesquisa qualitativa;
- **múltiplos métodos** – pesquisadores lançam mão de diversos tipos de dados, de modo a entenderem seu significado e a organizá-los em categorias ou temas;
- **raciocínio complexo por meio da lógica indutiva e dedutiva** – pesquisadores criam categorias ou temas “de baixo para cima”, dispondo os dados coletados a unidades de informação cada vez mais abstratas. Nesse processo indutivo, pesquisadores avançam e retrocedem em relação às categorias, uma vez que as mesmas, com frequência, são confrontadas com os dados coletados;

- **significados dos participantes** – o estudo qualitativo deve refletir as perspectivas dos participantes da pesquisa. Portanto, durante todo o processo de pesquisa, pesquisadores devem atentar-se quanto à captura dos significados atribuídos por esses participantes em relação ao fenômeno investigado;
- **projeto emergente** – o processo de pesquisa qualitativa é emergente, no sentido de que o planejamento inicial proposto para a pesquisa pode não ser rigidamente prescrito. Isso pode ocorrer em razão de fases do processo da pesquisa serem modificadas após a interação do pesquisador em campo;
- **reflexão** – pesquisadores transmitem sua interpretação aos dados coletados em determinado estudo, posicionando-se quanto aos resultados obtidos na pesquisa.

O método da teoria fundamentada foi adotado neste estudo. Sob perspectiva histórica, Strauss e Corbin (2008) explicam que essa metodologia foi concebida por dois sociólogos – Barney Glaser e Anselm Strauss – que divergiam entre si quanto à tradição filosófica e de pesquisa as quais vieram, mas que juntos trouxeram contribuições igualmente importantes. Nesse ponto, Strauss graduou-se na *University of Chicago*, que, por sua vez, é reconhecida por forte tradição na realização de pesquisas qualitativas. Em decorrência disso, sua formação acadêmica foi influenciada por muitos estudos interacionistas e pragmáticos, que contribuíram para a elaboração desse método principalmente em relação aos seguintes aspectos: 1) a necessidade do pesquisador ir a campo para descobrir a realidade; 2) a importância da teoria originada a partir de dados coletados na pesquisa; 3) a complexidade dos fenômenos e das ações humanas; 4) a relevante atuação de pessoas na resolução de problemas; 5) a compreensão na qual pessoas atribuem significados; 6) a percepção de que a atribuição de significados é definida por meio da interação; 7) a capacidade do pesquisador torna-se sensível à natureza reveladora dos fatos; 8) a inter-relação entre condições, ação e consequência. Por outro lado, Glaser graduou-se na *Columbia University*, uma instituição cuja tradição sociológica era a realização de estudos quantitativos, com ênfase na pesquisa empírica.

De acordo com Charmaz (2009), pesquisadores que aplicam a teoria fundamentada constroem os resultados de suas pesquisas a partir do conjunto de

interações que testemunham e vivenciam no decorrer desse processo. Sob o espectro construtivista, a autora sintetiza as principais características da teoria fundamentada, a saber:

- o processo de pesquisa é fluido e interativo;
- o problema de pesquisa aponta as alternativas metodológicas iniciais para a coleta de dados;
- os pesquisadores integram a pesquisa que realizam;
- a análise emergente pode ocasionar a adoção de várias técnicas de coleta de dados, sendo realizadas em vários locais;
- a essência da teoria fundamentada são os níveis sucessivos de abstração da análise, em nível comparativo;
- pesquisadores fazem análises interacionistas e interpretativas.

Considerando o exposto, Charmaz (2009) ainda apresenta uma definição do termo “teoria fundamentada”, localizada no glossário de sua obra intitulada: *A construção da teoria fundamentada: guia prático para análise qualitativa*. Tal definição sintetiza pontos fundamentais ao entendimento desse método, como a criação de esquemas conceituais de teorias e de categorias analíticas, e outros procedimentos de coleta e análise de dados, conforme as seguintes explicações:

Teoria fundamentada – um método de condução da pesquisa qualitativa que se concentra na criação de esquemas conceituais de teorias por meio da construção da análise indutiva a partir dos dados. Por essa razão, as categorias analíticas são diretamente “fundamentadas” nos dados. O método privilegia a análise e não a descrição, as categorias novas em vez de ideias préconcebidas e teorias existentes, e a coleta de dados sequencial sistematicamente focada em vez de amplas amostras iniciais. Esse método distingue-se dos demais uma vez que implica o comprometimento do pesquisador na análise dos dados durante a coleta desses dados, utilizamos essa análise dos dados para instruir e determinar uma nova coleta de dados. Desse modo, a distinção nítida entre a coleta de dados e as fases de análise da pesquisa tradicional é intencionalmente obscurecida nos estudos da teoria fundamentada (CHARMAZ, 2009, p. 252, *grifo do autor*).

Para a consecução do objetivo geral proposto, o estudo também lançou mão de revisão sistemática da literatura. De acordo com Paré e Kitsiou (2016), tal recurso metodológico compreende: a formulação de uma questão de revisão e o desenvolvimento de uma estratégia de busca para identificar estudos elegíveis com

base em critérios preestabelecidos; a realização da pesquisa em diversas bases de dados e fontes de informação; a seleção de estudos e a extração de dados; a análise qualitativa ou quantitativa de dados; a apresentação e a interpretação de resultados.

De maneira semelhante, Boren e Moxley (2015) se referem à revisão sistemática da literatura como um processo constituído por seis etapas, as quais incluem: 1) elaboração de uma pergunta de revisão; 2) pesquisa, recuperação e seleção de artigos relevantes; 3) avaliação de qualidade; 4) extração de dados; 5) análise e síntese de dados; 6) redação da revisão sistemática.

De acordo com Poupart (2008), a realização de entrevista como técnica de coleta de dados é relevante no contexto de estudos cujo foco é explorar as visões de mundo de atores sociais. À vista disso, a entrevista torna-se ponto de partida na compreensão de condutas sociais e da maneira pela qual participantes da pesquisa atribuem significados a fenômenos que os cercam. Em abordagem similar, Gaskell (2005) argumenta que o mundo social é construído por pessoas em sua vivência cotidiana. Logo, a entrevista auxilia o pesquisador a entender diferentes realidades sociais, de modo a contribuir para a análise aprofundada sobre determinado tópico de pesquisa.

Ao caracterizar a entrevista individual em profundidade, Duarte (2011) explica que essa técnica qualitativa é utilizada para coletar dados a partir de percepções e vivências de informantes. Em síntese, Duarte (2011, p. 62) explica que se trata de um “recurso metodológico que busca, com base em teorias e pressupostos definidos pelo investigador, recolher respostas a partir da experiência subjetiva de uma fonte, selecionada por deter informações que se deseja conhecer”. Além disso, pode ser um componente útil para a investigação de problemas complexos, visto que possibilita construções da realidade à luz da própria subjetividade humana.

Baseado no exposto, **este estudo conduziu entrevistas semiestruturadas para a coleta de dados**. Ao explicar sobre essa técnica, Boni e Quaresma (2005) destacam a possibilidade da combinação entre perguntas abertas e fechadas, previamente definidas, nas quais o informante discorre sobre o tema proposto em contexto semelhante a uma conversa informal, como pode ser observado nas declarações seguintes:

As entrevistas semi-estruturadas combinam perguntas abertas e fechadas, onde o informante tem a possibilidade de discorrer sobre o tema proposto. O pesquisador deve seguir um conjunto de questões previamente definidas, mas ele o faz em um contexto muito semelhante ao de uma conversa informal. O entrevistador deve ficar atento para dirigir, no momento que achar oportuno, a discussão para o assunto que o interessa fazendo perguntas adicionais para elucidar questões que não ficaram claras ou ajudar a recompor o contexto da entrevista, caso o informante tenha “fugido” ao tema ou tenha dificuldades com ele. Esse tipo de entrevista é muito utilizado quando se deseja delimitar o volume das informações, obtendo assim um direcionamento maior para o tema, intervindo a fim de que os objetivos sejam alcançados (BONI; QUARESMA, 2005, p. 75).

Os dados coletados por meio de entrevistas foram analisados conforme três processos analíticos descritos na literatura científica por Strauss e Corbin (2008): codificação aberta, codificação axial e codificação seletiva. Nesse contexto, o *software* acadêmico *Maxqda* (empregado para análise de dados qualitativos e métodos mistos de pesquisa) foi utilizado para auxiliar procedimentos relacionados à criação de categorias e subcategorias, em termos de suas propriedades e dimensões.

Segundo Strauss e Corbin (2008), a codificação aberta refere-se à busca de um pesquisador em revelar e nomear conceitos contidos em um texto, agrupando-os conforme similaridades de natureza ou de significado em conceitos abstratos denominados “categorias”. Essas, por sua vez, podem ser entendidas como conceitos originários de dados coletados utilizados para representar fenômenos.

Tais categorias se desenvolvem em termos de suas propriedades e dimensões. As primeiras consistem em atributos que caracterizam as categorias, atribuindo-lhes significado. Já as segundas, abrangem os contextos nos quais as propriedades gerais de uma categoria se diferenciam, especificando-a e trazendo variação à teoria construída. Além disso, subcategorias podem surgir no processo de análise dos dados, na medida em que sejam necessárias especificações adicionais sobre determinada categoria. Logo, podem ser entendidas como conceitos que pertencem a uma categoria.

É importante explicar que as dimensões apresentadas na codificação aberta deste trabalho referem-se aos contextos das propriedades pertencentes às categorias identificadas no estudo. Trata-se de uma maneira lógica adotada para dar sentido a um primeiro nível de organização conceitual aos dados analisados. Tal maneira foi idealizada com base nas práticas contidas no modelo conceitual da pesquisa

(fundamentado na literatura científica, principalmente em modelos do ciclo de vida de dados) e refletidas no roteiro de entrevista, sendo também muito útil na diferenciação de categorias.

Na codificação axial, o pesquisador ocupa-se em reagrupar os dados que foram categorizados durante o processo de codificação aberta. Categorias são relacionadas às suas subcategorias de modo a fornecer informações mais detalhadas sobre o fenômeno investigado. Strauss e Corbin (2008, p. 124) explicam que a codificação axial “é o ato de relacionar categorias com subcategorias ao longo das linhas de suas propriedades e suas dimensões. Ela examina como as categorias se cruzam e se associam”.

A codificação seletiva é definida por Strauss e Corbin (2008, p. 143) como “o processo de integrar e de refinar categorias”. Por um lado, a integração consiste no ato de decidir a categoria central, também conhecida por categoria básica. Essa, por sua vez, reduz em poucas palavras todos os produtos das análises realizadas durante a codificação aberta e axial. Ou seja, a categoria central é capaz de compilar categorias com a finalidade de compor um todo explanatório. Por outro lado, refinar a teoria consiste na busca do pesquisador pela consistência interna, por meio da identificação de falhas lógicas, da complementação de categorias mal desenvolvidas e da validação de esquemas. À vista disso, a teoria é validada a partir da comparação com os dados brutos ou de sua exposição aos participantes da pesquisa para obter suas opiniões a respeito do trabalho de análise realizado pelo pesquisador.

3.3 Procedimentos metodológicos

De modo a alcançar o objetivo de *propor diretrizes para políticas institucionais de gestão de dados abertos de pesquisa em universidades*, três objetivos específicos foram definidos. Esta seção detalhará os procedimentos metodológicos que foram adotados para a execução de cada um dos objetivos específicos, a partir da descrição das fontes de dados e das técnicas de coleta e análise de dados.

3.3.1 Identificar práticas de gestão de dados de pesquisa realizadas em diferentes áreas, com base na literatura científica

Com o propósito de identificar como o corpo teórico no âmbito da Ciência da Informação tem refletido práticas de gestão de dados de pesquisa em diferentes áreas, o estudo analisou artigos de periódicos científicos por meio do método de revisão sistemática da literatura.

Os artigos foram selecionados a partir de buscas na base de dados *Library and Information Science Abstracts* (LISA), em razão de três aspectos. O primeiro diz respeito a essa base de dados ser quantitativamente representativa em relação ao número de registros que possui. O segundo refere-se à base ser qualitativamente significativa em relação à produção científica indexada. O terceiro, por seu turno, está associado à base de dados ter ampla cobertura em assuntos de interesse à Ciência da Informação.

Criada em 1969, a base de dados LISA se propõe a veicular fontes de informações científicas relacionadas ao contexto de profissionais de bibliotecas e especialistas em informação. Indexa periódicos de mais de 40 países, em mais de 20 idiomas, que tratam sobre diversos assuntos, tais como gestão da informação e do conhecimento, Biblioteconomia e recuperação *online* da informação (PROQUEST⁶³).

A coleta de dados ocorreu no mês de abril de 2021, mediante busca avançada na base de dados mencionada. No campo da busca foi inserida a estratégia: `ti("Research data management") and mainsubject("Research data management") and ab("practices")`, que corresponde a buscas dos termos no campo “Título do Documento – TI”, “Assunto principal – mainsubject” e “Resumo – AB”, respectivamente. Além disso, a busca ateu-se à seleção de documentos revisados por especialistas, sendo artigos provenientes de periódicos acadêmicos. Não houve restrição de datas ou de idiomas dos documentos.

Após a busca, foram recuperados 14 artigos publicados entre 2015 e 2020, sendo treze no idioma inglês e um no espanhol. Os resumos desses artigos foram

critérios examinados com a finalidade de verificar se tratavam temática pertinente ao estudo proposto, ou seja, se de fato abordavam em seu conteúdo práticas de gestão de dados de pesquisa. Como todos os artigos atendiam ao critério estabelecido, nenhum deles foi excluído. Portanto, 14 artigos compuseram o *corpus* dos artigos analisados neste estudo, cujas descrições relativas ao título, autoria e ano de publicação são apresentadas no Quadro 9:

Quadro 9 - Artigos selecionados para o estudo

ID	Título	Autor	Ano
1	Research data management at a public university in Malawi: the role of “three hands”	Chawinga; Zinn	2020
2	ERDMAS: An exemplar-driven institutional research data management and analysis strategy	Bellgard	2020
3	Research Data Management (RDM) Capabilities at the University of Ghana, Legon	Avuglah; Underwood	2019
4	Cultural obstacles to research data management and sharing at TU Delft	Plomp et al.	2019
5	Three camps, one destination: the intersections of research data management, FAIR and Open	Higman; Bangert; Jones	2019
6	The collaborative creation of a Research Data Management Librarian Academy (RDMLA)	Shipman; Tang	2019
7	FAIR x FAIR. Una aproximación pragmática a la gestión de datos de investigación en modo FAIR	Alcalá; Anglada	2019
8	Developing research data management services and support for researchers: A mixed methods study	Perrier; Barnes	2018
9	Research Data Management Practices of Faculty Members	Piracha; Ameen	2018
10	Research data management in Turkey: perceptions and practices	Aydinoglu; Dogan; Taskin	2017
11	Moving a brick building: UK libraries coping with research data management as a ‘wicked’ problem	Cox; Pinfield; Smith	2016
12	Variability in academic research data management practices Implications for data services development from a faculty survey	Whitmire; Boock; Sutton	2015
13	Research data management and openness: The role of data sharing in developing institutional policies and practices	Higman; Pinfield	2015
14	Assessing Research Data Management Practices of Faculty at Carnegie Mellon University	Van Tuyl; Michalek	2015

Fonte: elaborado pela autora

Para a análise dos artigos indicados no Quadro 9, o estudo lançou mão do método da revisão sistemática da literatura. Esse método caracteriza-se, sobretudo, por ser uma estratégia adotada na pesquisa científica que segue protocolos preestabelecidos, e por atribuir logicidade a um determinado *corpus* documental, conforme explica Galvão e Ricarte (2019):

[A revisão sistemática] é uma modalidade de pesquisa, que segue protocolos específicos, e que busca entender e dar alguma logicidade a um grande corpus documental (...) Está focada no seu caráter de reprodutibilidade por outros pesquisadores, apresentando de forma explícita as bases de dados bibliográficos que foram consultadas, as estratégias de busca empregadas em cada base, o processo de seleção dos artigos científicos, os critérios de inclusão e exclusão dos artigos e o processo de análise de cada artigo (GALVÃO; RICARTE, 2019, p. 58-59).

Um formulário foi utilizado para a análise de cada artigo selecionado no estudo. Como pode-se observar no Quadro 10, esse formulário abrangeu três campos principais: 1) dados de identificação – inclui dados descritivos dos artigos analisados, como autor, título, ano e lugar de publicação; 2) caracterização da pesquisa – engloba aspectos como o objetivo da pesquisa, estratégia de investigação, técnica para a coleta de dados e descritores atribuídos pelo(s) autor(es); 3) identificação de práticas de gestão de dados de pesquisa – campo que possibilitou assinalar práticas identificadas na literatura científica e adicionar outras, conforme verificado nos artigos.

Quadro 10 - Formulário de análise para a identificação de práticas de gestão de dados de pesquisa realizadas em diferentes áreas

Dados de identificação	Autor	
	Título do artigo	
	Ano	
	Lugar de publicação	
Caracterização da pesquisa	Objetivo da pesquisa	
	Estratégia de investigação	<input type="checkbox"/> Qualitativa <input type="checkbox"/> Quantitativa <input type="checkbox"/> Mista
	Técnica para a coleta de dados	
	Descritores atribuídos pelo autor	
Identificação de práticas de gestão de dados de pesquisa	Descrição de práticas	<input type="checkbox"/> Planejar Descrição:
		<input type="checkbox"/> Coletar Descrição:
		<input type="checkbox"/> Processar Descrição:
		<input type="checkbox"/> Compartilhar Descrição:
		<input type="checkbox"/> Preservar Descrição:
		<input type="checkbox"/> Descobrir Descrição:
		<input type="checkbox"/> Reutilizar Descrição:
		<input type="checkbox"/> Monitoramento e controle Descrição:
		<input type="checkbox"/> Outra: _____ Descrição:
		Observações

Fonte: elaborado pela autora

A Figura 17 representa a síntese dos principais procedimentos adotados na revisão sistemática da literatura realizada neste trabalho, os quais se concentram na execução das seguintes etapas: 1) busca na base de dados LISA; 2) verificação de artigos recuperados; 3) seleção de artigos; 4) análise de artigos.

Figura 17 – Síntese dos procedimentos adotados na revisão sistemática da literatura



Fonte: elaborado pela autora

3.3.2 Analisar práticas de gestão de dados de pesquisa realizadas por pesquisadores em diferentes áreas

Tendo em vista a execução desse objetivo específico, a população estudada foi constituída por bolsistas de produtividade em pesquisa (nível PQ-2) do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) vinculados a universidades federais brasileiras. Os critérios determinantes para a definição da população referida neste estudo foram:

- a concessão de bolsas de produtividade é um indicador relevante de atividade de pesquisa, o que evidencia a manipulação rotineira de dados por parte de pesquisadores;
- um conjunto de critérios devem ser observados para a concessão de bolsa de produtividade a pesquisadores, conforme estabelecido no item 3.1 do anexo III da Resolução Normativa (RN) n° 28, de 18 de dezembro de 2015 do CNPq:

O pesquisador deverá:

- a) possuir o título de doutor ou perfil científico equivalente;
- b) ser brasileiro ou estrangeiro em situação regular no País;
- c) dedicar-se às atividades constantes de seu pedido de bolsa, e
- d) poderá ser aposentado, desde que mantenha atividades acadêmico-científicas oficialmente vinculadas a instituições de pesquisa e ensino (CNPQ, 2015).

- além disso, bolsistas de produtividade em pesquisa são continuamente avaliados com a finalidade de classificação, enquadramento e progressão, por categoria e nível. Para tanto, são analisados critérios que incluem:

- a) mérito científico do projeto;
- b) relevância, originalidade e repercussão da produção científica do candidato;
- c) formação de recursos humanos em nível de Pós-Graduação;
- d) contribuição científica, tecnológica e de inovação, incluindo patentes;
- e) coordenação ou participação em projetos e/ou redes de pesquisa;
- f) inserção internacional do proponente;
- g) participação como editor científico;
- h) participação em atividades de gestão científica e acadêmica (CNPQ, 2015).

Para fim de esclarecimento a respeito dos requisitos mínimos para enquadramento e classificação de bolsistas de produtividade em pesquisa, torna-se necessário recorrer à RN nº 28, de 18 de dezembro de 2015 do CNPq e apontar alguns de seus aspectos. De acordo com a norma, pesquisadores são enquadrados em 2 categorias – Pesquisador 1 e Pesquisador 2, e, dentro dessas, são classificados por níveis. A categoria Pesquisador 1 abrange bolsistas que possuem, no mínimo, 8 anos de doutorado por ocasião da efetivação da bolsa. Por outro lado, a categoria Pesquisador 2 inclui bolsistas que possuem, no mínimo, 3 anos de doutorado por ocasião da efetivação da bolsa.

O pesquisador pertencente à categoria 1 é enquadrado em quatro níveis: A; B; C ou D; de acordo com comparações entre seus pares e informações sobre sua carreira obtidas nos últimos 10 anos. Agrupam-se no nível A pesquisadores que demonstrem excelência continuada quanto à produção científica e à formação de recursos humanos, além de serem líderes de grupos de pesquisa reconhecidos. Geralmente, tais pesquisadores desempenham liderança estratégica no Brasil dentro da área de pesquisa em que atuam. Já os pesquisadores pertencentes aos níveis B e C, além dos critérios da participação na formação de recursos humanos e da

produção científica e tecnológica, são avaliados em relação a sua contribuição em grupos de pesquisa e programas de graduação e pós-graduação na instituição em que estão vinculados. Por outro lado, a categoria 2 não admite estratificação por níveis. A produtividade dos pesquisadores é avaliada com referência aos últimos 5 anos, relacionada, principalmente aos trabalhos publicados e às orientações acadêmicas realizadas (CNPQ, 2015).

Os participantes desta pesquisa foram selecionados dentre pesquisadores pertencentes à categoria 2. Isso, em razão de as avaliações de produtividade serem mais recentes, tendo em vista a análise do CNPq sobre trabalhos e orientações realizadas por pesquisadores nos últimos 5 anos. O estudo considerou apenas pesquisadores com bolsas de produtividade em pesquisa ativas, ou seja, não foram admitidos aqueles com bolsas suspensas ou canceladas.

Além disso, os participantes da pesquisa foram bolsistas de produtividade em pesquisa vinculados a universidades federais brasileiras. O motivo principal para o referido recorte é a representatividade nacional que essas instituições apresentam em termos de pesquisa, extensão e ensino, inclusive em relação ao recebimento de recursos financeiros oriundos do governo federal brasileiro.

Com o propósito de evidenciar as diferenças intrínsecas à comunidade científica em relação às áreas do conhecimento e demarcar com clareza os reflexos das diferenças disciplinares nas práticas de gestão de dados abertos de pesquisa em universidades, o estudo lançou mão de perspectivas de duas áreas distintas: Sociologia e Química. A secção realizada respaldou-se, sobretudo, na operacionalização do conceito “culturas epistêmicas” de Knorr-Cetina (1999), que aponta a existência de divisões culturais entre grupos de pesquisadores. Tais divisões são evidenciadas no que se refere à organização da pesquisa, às ocupações profissionais e aos sistemas de classificação do conhecimento.

Considerando os aspectos mencionados, a seleção da amostra ocorreu por meio de amostragem aleatória simples, sendo composta por bolsistas de produtividade em pesquisa - (nível PQ-2) do CNPq da área de Química e de Sociologia vinculados a universidades federais brasileiras. É importante arrazoar que o recorte amostral adotado foi embasado no entendimento de que tais disciplinas são representantes

de duas grandes áreas: *Exact and Natural Sciences* (ENS) e *Social Sciences and Humanities* (SSH), respectivamente.

Para o contexto desta pesquisa, considerou-se a perspectiva de um *continuum* entre disciplinas, na qual a Química encontra-se em um extremo correspondente a áreas mais rígidas, enquanto a Sociologia situa-se no outro extremo, constituído por áreas mais flexíveis. Acrescente-se, por oportuno, que não houve juízo de valor em relação aos fundamentos e princípios adotados pelas disciplinas referidas, nem quanto ao conjunto de práticas que realizam e transmitem dentro da comunidade científica.

Para identificar pesquisadores que atendiam aos critérios de participação estabelecidos nesta pesquisa, foram realizadas buscas em *site* do CNPq⁶⁴, no qual pode-se verificar os processos em curso na modalidade de Bolsa de Produtividade em Pesquisa por área do conhecimento. Assim, os procedimentos adotados em relação à identificação de pesquisadores de Química e de Sociologia, respectivamente, são descritos conforme as seguintes etapas:

Com relação a pesquisadores no campo da Química:

1. a opção “Química” foi selecionada no campo “Engenharias, Ciências Exatas e da Terra”. Como resultado, foi recuperada listagem contendo o nome, nível, data de início e término da bolsa, instituição a que pertence e situação da bolsa de 717 pesquisadores (PQ1- A, B, C, D; PQ-2 e PQ-SR). Essa listagem foi replicada em uma planilha do programa editor *Microsoft Excel* para facilitar a visualização dos dados obtidos;
2. na planilha contendo as informações dos pesquisadores foram aplicados dois filtros. O primeiro foi utilizado com o propósito de selecionar apenas pesquisadores do nível PQ-2, de modo a excluir pesquisadores das demais categorias e respectivos níveis. Dessa forma, pesquisadores correspondentes à categoria (PQ1- A, B, C, D) e pesquisadores Produtividade em Pesquisa Sênior (PQ-SR) foram excluídos, restando assim, 415 pesquisadores PQ-2.

A partir disso, foi possível aplicar o segundo filtro na planilha com a finalidade de selecionar apenas pesquisadores vinculados a universidades federais. Logo,

64 Disponível em: http://plsql1.cnpq.br/divulg/RESULTADO_PQ_102003.curso

restaram 259 pesquisadores vinculados a 39 instituições. A situação ativa do recebimento da bolsa pelo pesquisador é demonstrada pela informação “Em folha de pagamento”. Após análise desse campo específico, observou-se que nenhum pesquisador apresentava suspensão da bolsa. Sendo assim, confirmou-se que 259 pesquisadores identificados se encontravam com bolsas ativas, ou seja, sem qualquer suspensão ou cancelamento.

Com relação a pesquisadores no campo da Sociologia:

1. semelhantemente, a opção “Sociologia” foi selecionada no campo “Ciências Humanas e Sociais Aplicadas”. Como resultado, foi recuperada listagem incluindo 221 pesquisadores (PQ1- A, B, C, D; PQ-2 e PQ-SR), sendo essa replicada em uma planilha do programa editor *Microsoft Excel*;
2. na planilha referida foram aplicados dois filtros. O primeiro foi estabelecido para selecionar apenas pesquisadores do nível PQ-2, possibilitando a exclusão de pesquisadores correspondentes à categoria (PQ1- A, B, C, D) e pertencentes à categoria Produtividade em Pesquisa Sênior (PQ-SR). Desse modo, restaram 121 pesquisadores PQ-2. O segundo filtro na planilha foi estabelecido com o propósito de selecionar apenas pesquisadores pertencentes a universidades federais. Assim, restaram 80 pesquisadores pertencentes a 24 instituições;
3. após análise do campo “situação” da bolsa, verificou-se que um pesquisador apresentou suspensão da bolsa. Assim, 79 pesquisadores identificados permanecem com bolsas ativas, o que significa não apresentarem qualquer suspensão ou cancelamento.

Separadamente, houve um sorteio dentre os 259 pesquisadores em Química e outro dentre os 79 pesquisadores em Sociologia, com a finalidade de selecionar de modo aleatório os prováveis participantes da pesquisa e ordená-los sequencialmente. Respeitada essa ordem, os pesquisadores de cada uma das áreas estudadas (Química e Sociologia) que atendiam aos critérios anteriormente declarados foram convidados via e-mail (disponibilizado publicamente na aba de contato do Currículo Lattes, e/ou contato disponível em trabalhos científicos publicados) a participarem de entrevistas semiestruturadas realizadas por videoconferência.

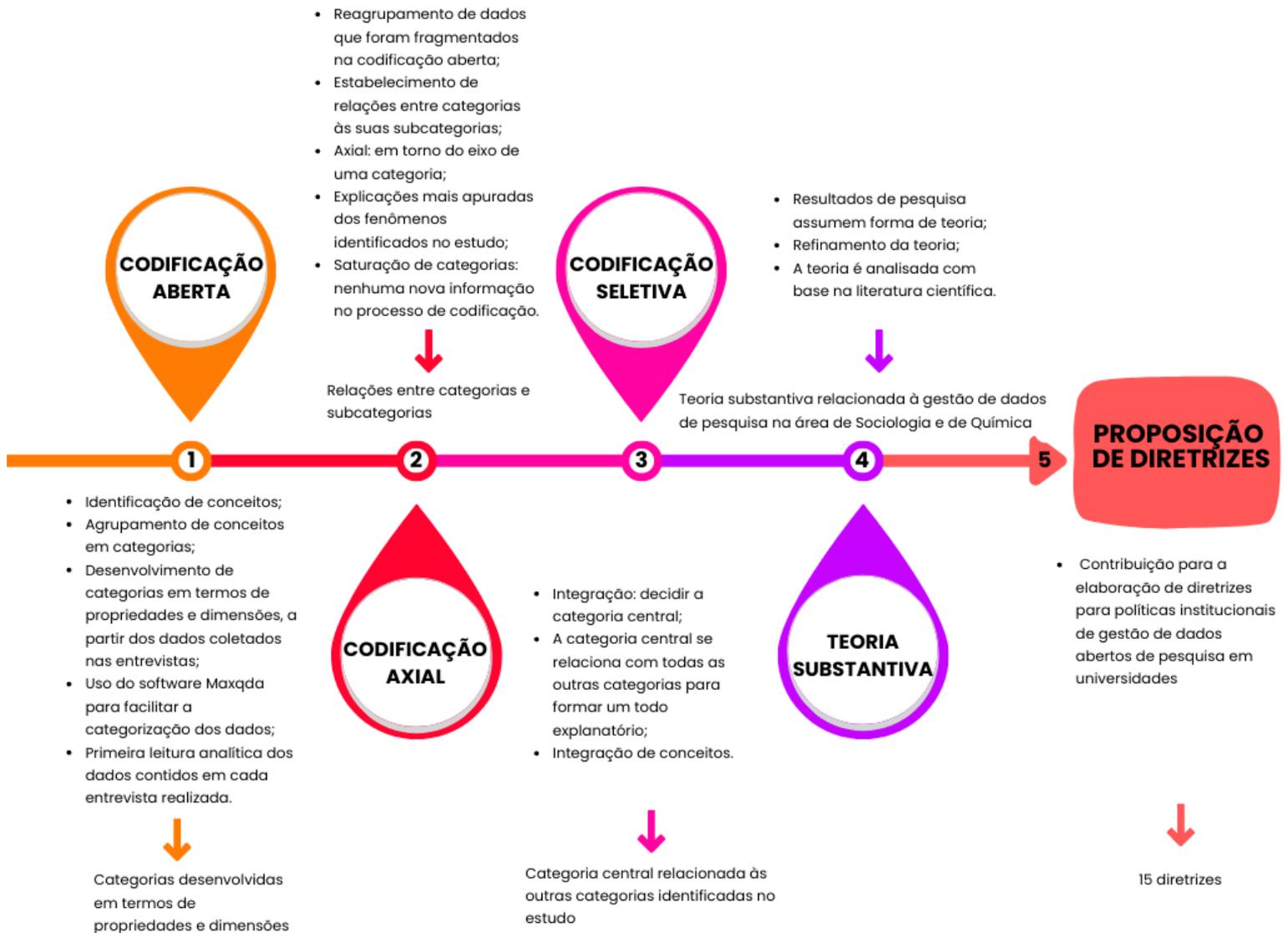
Foram realizadas dez entrevistas com pesquisadores de Sociologia e dez com pesquisadores de Química, entre 6 de outubro de 2021 e 29 de abril de 2022, com duração média de uma hora cada uma. Informa-se que os dados foram coletados até não serem mais discernidos outros dados relevantes a serem acrescentados às categorias identificadas na pesquisa. Por essa razão, considera-se que o trabalho atingiu o ponto de saturação amostral.

A análise dos dados coletados em cada uma das entrevistas procedeu conforme a aplicação do método da teoria fundamentada, o que inclui os processos analíticos de codificação aberta, codificação axial, codificação seletiva e elaboração da teoria substantiva nas duas áreas consideradas no presente estudo – Sociologia e Química.

Por se tratar de processos analíticos complexos, a Figura 18 é relevante nesse contexto em razão do detalhamento que apresenta sobre as etapas metodológicas e os resultados obtidos em cada uma das mesmas. Convém ressaltar que tais etapas foram baseadas nos ensinamentos apregoados por Strauss e Corbin (2008), no que se refere a técnicas para o desenvolvimento de teoria fundamentada. De modo resumido, pode-se apontar os seguintes aspectos relacionados aos processos analíticos realizados no presente estudo:

- a codificação aberta possibilitou o agrupamento de conceitos em categorias, as quais foram desenvolvidas em termos de propriedades e dimensões;
- a codificação axial relacionou categorias e subcategorias identificadas;
- a codificação seletiva ocorreu mediante a integração de conceitos. Sendo assim, foi possível identificar a categoria central e suas relações com as outras categorias do estudo;
- os resultados da pesquisa tomaram forma de uma teoria substantiva relacionada à gestão de dados de pesquisa na área de Sociologia e de Química, sendo também analisada à luz da literatura científica.

Figura 18 – Procedimentos de análise de dados relacionados às práticas de gestão de dados de pesquisa realizadas por pesquisadores em diferentes áreas



Fonte: elaborado pela autora

3.3.3 Descrever fatores que influenciam a abertura de dados de pesquisa, segundo pesquisadores em diferentes áreas

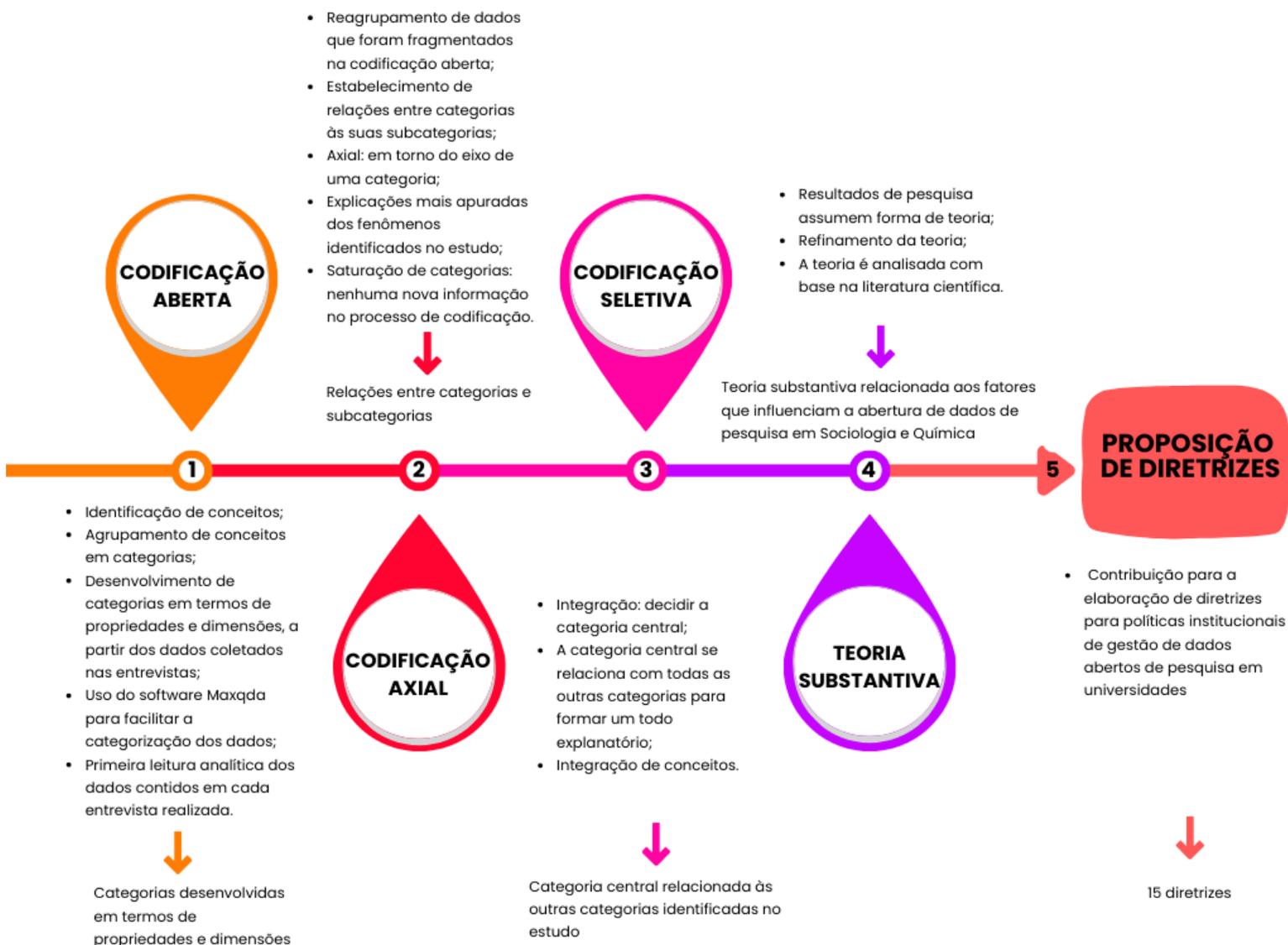
Para atingir esse objetivo específico, a amostra da pesquisa foi a mesma mencionada na Seção 3.3.2 deste trabalho. Trata-se de bolsistas de produtividade em pesquisa - (nível PQ-2) do CNPq da área de Sociologia e de Química vinculados a universidades federais brasileiras.

Semelhantemente aos procedimentos metodológicos adotados para a consecução do objetivo específico 2 – analisar práticas de gestão de dados de pesquisa realizadas por pesquisadores em diferentes áreas – os dados foram coletados com os mesmos entrevistados e roteiro de entrevista, simultaneamente. A diferença reside apenas no foco da análise realizada. Para o objetivo 2, a atenção manteve-se na identificação de conceitos relacionados à gestão de dados de pesquisa. Enquanto no objetivo 3, o principal conceito abordado foi o da abertura de dados de pesquisa.

A análise dos dados coletados em cada uma das entrevistas procedeu conforme os mesmos processos realizados para a análise de dados associados a práticas de gestão de dados de pesquisa realizadas por pesquisadores em diferentes áreas. Isso porque o estudo também lançou mão de teoria fundamentada para atingir o objetivo de descrever fatores que influenciam a abertura de dados de pesquisa em diferentes áreas.

Trata-se da realização dos processos analíticos: codificação aberta; codificação axial; codificação seletiva e elaboração da teoria substantiva nas duas áreas estudadas neste trabalho – Sociologia e Química. Nessa direção, a Figura 19 apresenta os procedimentos de análise de dados relacionados aos fatores que influenciam a abertura de dados de pesquisa em diferentes áreas.

Figura 19 – Procedimentos de análise de dados relacionados aos fatores que influenciam a abertura de dados de pesquisa em diferentes áreas



Fonte: elaborado pela autora

3.3.4 Síntese dos procedimentos metodológicos adotados para cada objetivo específico proposto no trabalho

Considerando a totalidade dos procedimentos metodológicos realizados com vista ao cumprimento dos objetivos específicos declarados nesta pesquisa (descritos nas Seções 3.3.1 a 3.3.3 do presente trabalho), o Quadro 11 apresenta uma síntese de cada um desses procedimentos em relação à população, amostra, fonte dos dados, técnica de coleta e análise de dados do estudo.

Quadro 11 – Procedimentos metodológicos da pesquisa

OBJETIVO ESPECÍFICO	POPULAÇÃO	AMOSTRA	FONTE	TÉCNICA DE COLETA	TÉCNICA DE ANÁLISE
Identificar práticas de gestão de dados de pesquisa realizadas em diferentes áreas, com base na literatura científica	Literatura científica em Ciência da Informação	Artigos sistematicamente selecionados	Base de dados LISA	Pesquisa bibliográfica	Revisão sistemática da literatura
Analisar práticas de gestão de dados de pesquisa realizadas por pesquisadores em diferentes áreas	Bolsistas de produtividade em pesquisa (PQ-2) vinculados a universidades federais brasileiras	Bolsistas de produtividade em pesquisa (PQ-2) na área de Sociologia e de Química vinculados a universidades federais brasileiras	CNPq	Entrevista	Codificação aberta, codificação axial e codificação seletiva
Descrever fatores que influenciam a abertura de dados de pesquisa, segundo pesquisadores em diferentes áreas					

Fonte: elaborado pela autora

3.3.5 Descrição do pré-teste

O estudo lançou mão de um pré-teste com a finalidade de mitigar inconsistências na pesquisa, principalmente em relação à coleta e análise de dados. Esse recurso foi fundamental para implementar melhorias no roteiro de entrevista, eleger um *software* para auxiliar na codificação de dados e avaliar o modo pelo qual as entrevistas deveriam ser conduzidas, o que inclui a observação do tempo de realização e até mesmo da forma de se fazer perguntas ao entrevistado, em um contexto de videoconferência.

Assim sendo, o pré-teste foi realizado no mês de agosto de 2021 com um pesquisador convidado via e-mail a participar de entrevista semiestruturada realizada por videoconferência. Tal participante preencheu todos os critérios adotados para a seleção da amostra no estudo, ou seja, tratava-se de bolsista de produtividade em pesquisa (nível PQ-2) na área de Sociologia do CNPq vinculado a uma universidade federal brasileira.

Com o pré-teste foi possível identificar perguntas que pareciam não ser suficientemente claras ao entrevistado, gerando margens para interpretações desacertadas e fora do contexto os quais foram estabelecidas. Isso foi percebido pela entrevistadora, na medida em que teve que repetir tais perguntas e explicar o sentido correspondente ao que se almejava. Nesse sentido, questões foram complementadas, adicionando termos explicativos ou retirando-os para evitar duplicidade de sentidos.

A entrevista foi gravada e os dados obtidos foram transcritos para posterior análise. O pré-teste também foi importante para a escolha da plataforma digital em que as entrevistas seriam realizadas. Assim, verificou-se que a plataforma *Zoom* seria a mais adequada para o propósito da pesquisa, principalmente por três razões: 1) possibilitar a gravação de áudio e vídeo da reunião, sem a necessidade de pagamento; 2) viabilizar reunião por tempo suficiente para a condução das entrevistas; 3) estar integrada ao recurso do Google agenda.

Os dados degravados foram importados para o *software Maxqda*, permitindo a criação de categorias, subcategorias, descrição de propriedades e organização do conteúdo textual, por meio da compilação e apresentação de diferentes textos integrantes de uma mesma categoria e a exploração de recursos visuais que facilitam a compreensão dos dados manuseados.

A entrevista do pré-teste durou 54 minutos. Observou-se que esse tempo foi suficiente para acolher todo o propósito das perguntas contidas no roteiro de entrevista, não sendo necessário a redução no número de questões. Além disso, a última pergunta do roteiro abordou sugestões/comentários a respeito de aspectos da entrevista, como o roteiro utilizado, o modo pelo qual a entrevistadora conduziu as perguntas e a dinâmica adotada. O feedback obtido foi otimista em relação à adequação da entrevista ao estudo proposto. Por fim, informa-se que os dados

obtidos no pré-teste realizado não foram incluídos neste trabalho.

3.3.6 Considerações éticas

Tendo como pano de fundo aspecto disciplinado pelo Conselho Nacional de Saúde, por meio da Resolução 466, de 12 de dezembro de 2012 (CNS, 2012) e da Resolução 510, de 7 de abril de 2016 (CNS, 2016) (sendo a última um expoente dos princípios éticos das pesquisas em ciências humanas e sociais) quanto à participação de seres humanos em pesquisas científicas, informa-se que o presente estudo foi submetido e aprovado no Comitê de Ética em Pesquisa em Ciências Humanas e Sociais (CEP/CHS) da Universidade de Brasília, em 2021.

Os riscos, apesar de mínimos, podem estar associados a exposições de vivências de pesquisadores em seus relatos sobre as práticas de gestão de dados das pesquisas que conduzem, ainda que as identidades desses participantes sejam mantidas sob sigilo. Ademais, destaca-se que a participação nesta pesquisa foi voluntária e livre de qualquer remuneração ou benefício direto. Benefícios indiretos podem ser alcançados, no sentido de que os achados da pesquisa poderão contribuir, em nível prático, para a gestão de dados abertos de pesquisa em diferentes áreas no contexto de universidades.

Sob orientação do CEP/CHS, foi solicitada a assinatura de Termo de Consentimento Livre e Esclarecido (TCLE) a todos os participantes da pesquisa, com a finalidade de mantê-los conscientes a respeito dos objetivos, uso e disseminação das informações obtidas, além de documentar a livre vontade em participar do estudo.

Assim também, lhes foi solicitada anuência em relação ao Termo de autorização para utilização de imagem e som de voz para fins de pesquisa, ainda que a imagem não seja disponibilizada publicamente em qualquer meio ou recurso, ou que os sons de voz obtidos na pesquisa sejam utilizados apenas para fim de gravação de áudio, viabilizando posterior análise de seu conteúdo. Ressalta-se, ainda, que foi atribuído nome fictício aos entrevistados para que suas identidades não sejam reveladas, e que todas as informações coletadas na pesquisa serão utilizadas e disseminadas, exclusivamente, para propósitos acadêmicos e de pesquisa científica.

4 APRESENTAÇÃO E DISCUSSÃO DOS RESULTADOS

Este capítulo dedica-se à apresentação e discussão dos resultados obtidos no estudo, conforme os três objetivos específicos propostos inicialmente, os quais serão abordados na seguinte ordem:

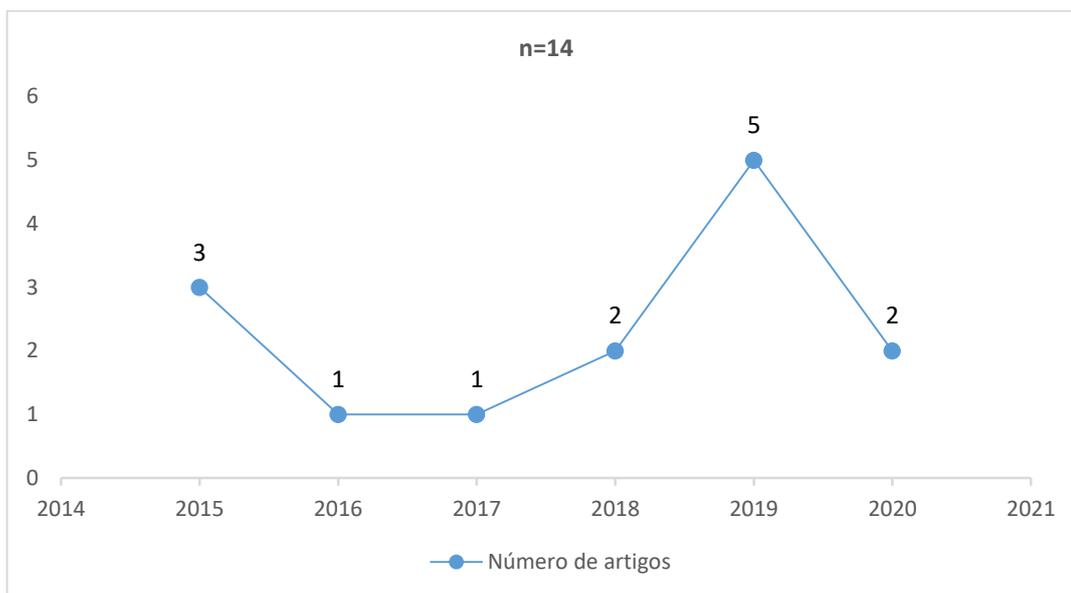
1. identificar práticas de gestão de dados de pesquisa em diferentes áreas, com base na literatura científica;
2. analisar práticas de gestão de dados de pesquisa realizadas por pesquisadores em diferentes áreas – primeiramente, serão apresentados os resultados referentes à área de Sociologia em cada um dos processos da teoria fundamentada (codificação aberta, codificação axial, codificação seletiva e teoria substantiva). Logo após, os resultados relativos à área de Química serão expostos, conforme os mesmos processos mencionados;
3. descrever fatores que influenciam a abertura de dados de pesquisa, segundo pesquisadores em diferentes áreas – tal como foi descrita a estratégia de apresentação dos resultados relacionados às práticas de gestão de dados de pesquisa realizadas por pesquisadores em diferentes áreas, primeiramente, serão apresentados os resultados referentes à área de Sociologia em cada um dos processos da teoria fundamentada (codificação aberta, codificação axial, codificação seletiva e teoria substantiva), e logo depois, os relativos à Química.

4.1 Identificar práticas de gestão de dados de pesquisa em diferentes áreas, com base na literatura científica

A análise de artigos recuperados na base de dados LISA, por meio de revisão sistemática da literatura, possibilitou a identificação de doze práticas de gestão de dados de pesquisa realizadas em diferentes áreas. Como ocorrem, em qual contexto estão inseridas e sob qual dinâmica se estabelecem tais práticas, foram aspectos levantados no estudo, e que, posteriormente, serão abordados em detalhes na Seção 4.1.5. Isso porque considerou-se a necessidade de, primeiramente, caracterizar a literatura investigada nos seguintes termos: número de artigos publicados por ano; lugar de publicação; tipo de autoria; palavras-chave atribuídas pelo autor. Dessa forma, as seções 4.1.1 a 4.1.4 tratam sobre tal caracterização.

4.1.1 Número de artigos publicados por ano

Os 14 artigos analisados nesta etapa do estudo datam entre 2015 e 2020. O Gráfico 1 apresenta o número desses artigos publicados por ano. Nele observa-se um aumento no número de artigos publicados entre 2017 e 2019. Nesse último ano, foram contabilizados cinco artigos – o maior número no período abordado. Apesar disso, em 2020, percebe-se um decréscimo nas quantidades observadas, sendo registrados dois artigos nesse ano. Contudo, infere-se que tal decréscimo pode ter ocorrido devido a artigos ainda não terem sido agregados à base de dados LISA no momento em que a coleta de dados para este trabalho foi realizada. Portanto, não há informação suficiente para apontar qualquer tendência de declínio.

Gráfico 1 – Número de artigos publicados por ano (2015-2020)

Fonte: elaborado pela autora

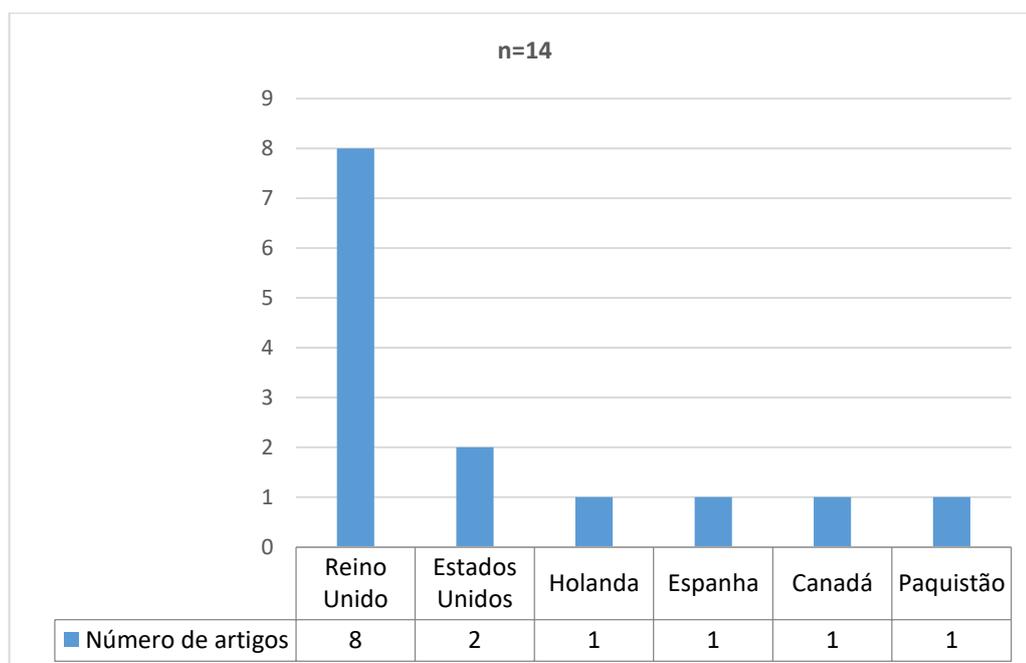
Com base no Gráfico 1, toma-se nota de que o tópico da gestão de dados de pesquisa tem sido presente na literatura científica publicada nos últimos anos. Todavia, como verificado por Guimarães e Bezerra (2019), discussões a esse respeito apontaram no corpo teórico desde pelo menos a década de 1960. Os autores realizaram pesquisa que empregou técnicas de bibliometria com o objetivo de identificar as principais características da produção em periódicos científicos sobre gestão de dados. Para tanto, foi realizado um levantamento nas bases de dados *Web of Science*, *SCOPUS*, *ScienceDirect* e *Emerald*, recuperando assim, 3.683 artigos científicos, publicados entre os anos 1963 e 2018.

Dentre os principais resultados obtidos no estudo de Guimarães e Bezerra (2019), destaca-se que o artigo recuperado mais antigo do conjunto analisado foi publicado em 1963, intitulado: “*Modern Data Management of Medical Board Reports*”, de autoria de Hall. Nas décadas seguintes, observou-se que a temática foi intensificada na literatura, contabilizando o maior número de artigos em 2017, com 210 registros. Assim também, verificou-se uma forte tendência de crescimento, uma vez que os artigos publicados entre 2013 e 2017 já representavam 25,58% do total de artigos integrantes no *corpus* analisado.

4.1.2 Lugar de publicação

O lugar de publicação dos artigos examinados nesta pesquisa também foi alvo de análise. Nesse sentido, o Gráfico 2 associa o número de artigos desse conjunto e o lugar onde foram publicados. Nele observa-se que o Reino Unido (com oito artigos), seguido dos Estados Unidos (com dois artigos) concentram o maior número de publicações. Juntos, somam dez artigos, o que corresponde a mais da metade do total do *corpus* analisado. Holanda, Espanha, Canadá e Paquistão contabilizaram a publicação de um artigo cada.

Gráfico 2 – Lugar de publicação dos artigos analisados



Fonte: elaborado pela autora

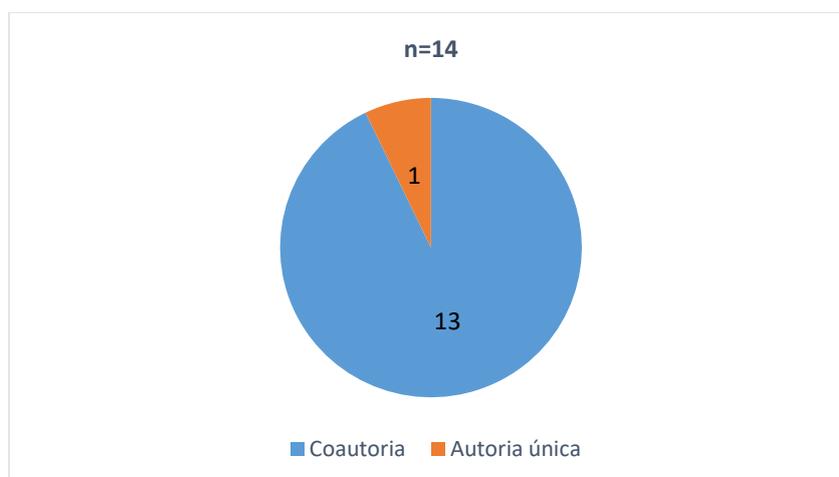
Considerando a concentração de artigos publicados no Reino Unido e Estados Unidos demonstrada no Gráfico 2, nota-se que tal ocorrência se refere a Estados desenvolvidos, ou países do Norte, como também foi evidenciado por Costa (2017). Propondo-se a caracterizar, a partir da literatura científica, fatores que influenciam a comunicação de dados de pesquisa, a autora observou o predomínio de países do Norte em relação a esse tópico específico. No que se refere ao lugar de origem do

primeiro autor de cada artigo analisado em sua pesquisa (no total de 13 artigos), Costa (2017) notou que o Reino Unido e os Estados Unidos foram predominantes. De treze autores, constatou-se que seis eram do Reino Unido e cinco dos Estados Unidos.

Tomando o Reino Unido como exemplo de uma região que se destaca no cenário internacional devido a iniciativas relacionadas à gestão de dados de pesquisa, pode-se mencionar a *expertise* do *UK Data Archive*⁶⁵. Desde 1967, a *University of Essex* abriga o *UK Data Archive*, ainda enquanto era conhecido por *Social Science Research Council Data Bank*. Em 2005, a iniciativa tornou-se local de armazenamento de registros públicos para o Arquivo Nacional. Atualmente, o *UK Data Archive* fornece suporte e treinamento a pesquisadores, além de acesso à coleção de dados sociais e econômicos do Reino Unido. Ademais, integra a Rede Internacional de Acesso a Dados, facilitando o uso de dados científicos entre centros de pesquisa na França, Alemanha, Holanda e Reino Unido.

4.1.3 Tipo de autoria

Com relação ao tipo de autoria dos artigos analisados no estudo, o Gráfico 3 associa o número de artigos por tipo de autoria – trabalhos em autoria única ou em coautoria. Tal dado é relevante, na medida em que apesar de não ser sinônimo de colaboração científica, aponta fortes indícios nessa direção. Assim, dentre os artigos investigados, o tópico da gestão de dados de pesquisa tem sido predominantemente discutido na literatura científica em nível de coautoria. Apenas um artigo foi publicado em autoria única.

Gráfico 3 – Número de artigos publicados em autoria única e em coautoria

Fonte: elaborado pela autora

Vanz e Stumpf (2010) esclarecem que a ocorrência de coautoria em publicações tem sido utilizada nas áreas de bibliometria e cientometria para examinar o fenômeno da colaboração entre pesquisadores, instituições e países. As autoras ainda explicam que, com frequência, ambos os termos são entendidos como sinônimos. Porém, é pertinente ater-se ao fato de que a coautoria constitui apenas um aspecto da colaboração científica, pois não a mensura em sua totalidade. Isso porque nem toda colaboração implica em artigo, e nem todo trabalho em coautoria é resultado de colaboração.

4.1.4 Palavras-chave atribuída pelo autor

As palavras-chave atribuídas pelos autores dos artigos investigados no estudo foram analisadas com a finalidade de identificar os principais assuntos discutidos neste *corpus*. Isso porque considera-se que tais palavras podem representar conceitos relevantes no campo da gestão de dados de pesquisa. Assim, a Tabela 2 apresenta as palavras-chave atribuídas pelos autores dos referidos artigos, o número de ocorrência de cada palavra e sua representatividade percentual.

Tabela 2 – Palavras-chave atribuídas pelos autores

Palavras-chave	Ocorrência	Porcentagem (%)
Research data management/RDM	12	16,90
Data sharing	4	5,63
Research data/Datos de investigación	3	4,22
Librarians	2	2,82
Open data	2	2,82
Open Science/Ciência aberta	2	2,82
Principios fair/FAIR	2	2,82
Data repository/Repositorios de datos	2	2,82
Academic libraries	2	2,82
Research data services	2	2,82
Palavras com apenas 1 ocorrência	38	53,52
Total	71	100

Fonte: elaborado pela autora

Conforme mostra a Tabela 2, das 71 palavras-chave identificadas, as mais utilizadas nos artigos foram: “Research data management” (gestão de dados de pesquisa, em português), representando 16,90%, seguida de “data sharing” (compartilhamento de dados, em português), com 5,63%, e de “research data” (dados de pesquisa, em português), com 4,22%. Outras palavras-chave também foram utilizadas, porém, com menor frequência (2,82%), como é o caso das palavras-chave: “librarians” (bibliotecários, em português); “open data” (dados abertos, em português); e “open science” (ciência aberta, em português). Toma-se nota, ainda, que palavras-chave com apenas uma ocorrência corresponderam a mais de 53% das 71 palavras-chave identificadas em todo o conjunto de artigos. Tal informação oferece indícios de que a gestão de dados de pesquisa é um tópico que está relacionado a diversos outros assuntos e temas da literatura científica.

4.1.5 Práticas identificadas

No estudo foram identificadas doze práticas de gestão de dados de pesquisa realizadas em diferentes áreas, sendo essas: 1) planejamento; 2) coleta; 3) processamento técnico; 4) compartilhamento; 5) preservação; 6) descoberta; 7) reuso;

8) monitoramento e controle; 9) elaboração de políticas para a gestão de dados de pesquisa; 10) suporte à análise de dados; 11) treinamento sobre a gestão de dados de pesquisa orientado a profissionais de informação; 12) treinamento sobre a gestão de dados de pesquisa orientado a pesquisadores. Tais práticas são mencionadas na literatura científica da seguinte maneira:

4.1.5.1 Planejamento da gestão de dados de pesquisa

Dentre os artigos analisados no estudo, quatro abordaram a prática do planejamento da gestão de dados de pesquisa, os quais são apresentados no Quadro 12, e, posteriormente, serão explorados ao longo da narrativa.

Quadro 12 – Artigos que abordaram o planejamento da gestão de dados de pesquisa

Prática	Título	Autor(es)	Ano
Planejar	Cultural obstacles to research data management and sharing at TU Delft	PLOMP et al.	2019
	FAIR x FAIR. Una aproximación pragmática a la gestión de datos de investigación en modo FAIR	ALCALÁ; ANGLADA	2019
	Research data management and openness: The role of data sharing in developing institutional policies and practices	HIGMAN; PINFIELD	2015
	Assessing Research Data Management Practices of Faculty at Carnegie Mellon University	VAN TUYL; MICHALEK	2015

Fonte: elaborado pela autora

Considerando o estudo de Plompe *et al.* (2019), que se concentrou em apresentar o projeto intitulado *Data Stewardship* desenvolvido no âmbito da *Delft University of Technology* (TU Delft), observa-se que a prática do planejamento da gestão de dados de pesquisa tem ocorrido entre pesquisadores da Universidade. Esse projeto abrange melhorias nas atuais práticas de gestão e compartilhamento de dados, de modo a introduzir mudanças relevantes relacionadas ao corpo docente e a oferecer apoio a pesquisadores nesse sentido.

Por meio da aplicação de questionários e realização de entrevistas informais em seis das oito faculdades da TU Delft, entre 2017 e 2018, Plompe *et al.* (2019) verificaram que a biblioteca da TU Delft tem apoiado pesquisadores, no que diz

respeito à elaboração do Plano de gestão de dados. Nessa direção, pesquisadores têm acesso ao *DMPonline*, uma plataforma *online* utilizada para a criação de plano de gestão de dados e que fornece diversos modelos para uso.

A TU Delft é considerada a maior universidade técnica da Holanda, com cerca de 5.000 funcionários e 23.500 estudantes. Dentre suas principais características, destaca-se o foco em estudos técnicos que coletam grandes quantidades de dados numéricos oriundos de experimentos físicos e simulações em computador, e o uso de diversas ferramentas de *software* para o processamento de dados obtidos. Entretanto, apesar de amplo foco na realização de pesquisas no campo da engenharia e ciências técnicas, as faculdades da TU Delft apresentam diferenças significativas relativas a metodologias aplicadas e tipos de dados produzidos. Como exemplo, pode-se mencionar que na Faculdade de Engenharia Elétrica, Matemática e Ciência da Computação, a maioria dos projetos de pesquisa realizados possui grande componente computacional, enquanto na Faculdade de Tecnologia, Política e Gestão, pesquisadores frequentemente obtêm dados de fontes pessoais derivadas de pesquisas quantitativas e qualitativas (PLOMPE et al., 2019).

Tendo como pano de fundo os esforços de universidades da Catalunha na prestação de serviços de apoio à investigação científica – o que inclui a criação do *Consorti de Serveis Universitaris de Catalunya* (CSUC) – uma comissão integrada por vice-reitores de pesquisa solicitou um relatório acerca da infraestrutura apontada pela Comissão Europeia (CE) necessária para a publicação de dados de pesquisa, conforme os princípios FAIR (*Findable, Accessible, Interoperable, Reusable*). Esse relatório foi concluído em 2019, e apresenta dados obtidos por meio de análise documental e de entrevistas realizadas com especialistas de instituições de ensino. Os principais pontos dessa publicação foram abordados em estudo de Alcalá e Anglada (2019).

No que concerne à prática do planejamento da gestão de dados, destaca-se a criação de um instrumento que possibilita elaborar, compartilhar e exportar um Plano de gestão de dados, de acordo com os requisitos estabelecidos pela CE. Tal instrumento (eiNa DMP) faz uso de *software* desenvolvido pelo *Digital Curation Center* (DCC) e *University of California Curation Center* (UC3). Além disso, apresenta descrições e exemplos a respeito de cada aspecto que o plano deve incluir. Trata-se

de uma espécie de guia mantido sob colaboração entre diferentes universidades membros do CSUC, podendo ser adaptado para instituição ou grupo de usuários específicos.

Higman e Pinfield (2015) conduziram estudo cujo objetivo foi investigar o que está impulsionando a criação de políticas e a realização de práticas de gestão de dados de pesquisa em Instituições de Ensino Superior (IES) do Reino Unido e a função do compartilhamento de dados nesse processo. Para tanto, duas estratégias de coleta e análise de dados foram adotadas sequencialmente na pesquisa. Por um lado, 37 políticas de gestão de dados disponíveis nas IESs do Reino Unido foram examinadas. Por outro lado, foram realizadas 11 entrevistas em duas IES com funcionários que prestam serviços de gestão de dados de pesquisa, o que inclui bibliotecários, equipes de escritório de pesquisa e profissionais da área de Tecnologia da Informação. Os dados das entrevistas foram codificados e analisados com o auxílio do *software* NVivo.

Dentre os resultados obtidos no estudo de Higman e Pinfield (2015), destaca-se que planos de gestão de dados têm sido amplamente requisitados por financiadores de pesquisas, e que o planejamento da gestão de dados foi a etapa mais mencionada do ciclo de vida dos dados de pesquisa nos documentos de políticas analisados, conforme apresentado na Tabela 3:

Tabela 3 – O número e a proporção de documentos de política que mencionaram os principais estágios do ciclo de vida dos dados de pesquisa

Estágios do ciclo de vida dos dados de pesquisa	Número de políticas	%
Planos de gestão de dados	37	100
Gestão ativa de dados	17	46
Disposição	14	38
Preservação	36	97
Nota n= 37		

Fonte: traduzido e adaptado de Higman e Pinfield (2015, p. 370)

Semelhantemente, Van Tuyl e Michalek (2015) observaram em seus estudos que os denominados Planos formais de gestão de dados eram extensamente utilizados por pesquisadores, devido a solicitações de agências de financiamento. Nesse estudo, os autores tinham o objetivo de compreender as necessidades de serviços de dados de pesquisadores da *Carnegie Mellon University* (CMU).

Para tanto, Van Tuyl e Michalek (2015) lançaram mão de um *survey* conduzido em uma amostra do corpo docente da universidade para analisar as tendências gerais na gestão de dados. Tal amostra incluiu docentes de diferentes áreas, como Ciências Humanas e Sociais, Ciências, Engenharia, Negócios, Ciência da Computação, Artes e Políticas. Assim também, os autores conduziram entrevistas com uma amostra reduzida do corpo docente, com o propósito de analisar, em profundidade, as práticas adotadas em relação aos serviços de dados nos *campus*. Tome-se nota, ainda, de que informações foram obtidas a partir da administração de pesquisa da universidade para auxiliar a contextualização de alguns resultados.

No que se refere, especificamente, à elaboração de Planos de gestão de dados, destaca-se que muitos docentes expressaram estar frustrados com a criação de planos informais, ou de padrões de planejamento utilizados somente em nível de laboratório ou do projeto nos quais atuam. Isso porque, geralmente, tais planos não tratavam de diretrizes para a gestão de dados, mas apenas de expectativas gerais a respeito de como dados devem ser gerenciados. Logo, ofereciam uma vaga estrutura que não era tão significativa para conduzir processos relacionados à gestão de dados.

4.1.5.2 Coleta de dados

No que se refere à prática de coleta de dados associada ao contexto da gestão de dados de pesquisa, um artigo foi identificado acerca desse tópico específico, conforme apresentado no Quadro 13.

Quadro 13 – Artigo que abordou a coleta de dados no contexto da gestão de dados de pesquisa

Prática	Título	Autor(es)	Ano
Coletar	Variability in academic research data management practices: Implications for data services development from a faculty survey	WHITMIRE; BOOCK; SUTTON	2015

Fonte: elaborado pela autora

O artigo indicado no Quadro 13 trata de estudo de caso realizado em 2015, por Whitmire, Boock e Sutton, que analisou o desenvolvimento de serviços de dados

na *Oregon State University* (OSU). Assim, um questionário foi distribuído por *e-mail* para o corpo docente da universidade, incluindo professores associados, professores-assistentes e pesquisadores de pós-doutorado atuantes em diversas áreas, como Ciências agrícolas, Educação, Saúde pública, Ciências humanas, Negócios e Silvicultura. As perguntas da pesquisa abordaram aspectos relacionados à gestão de dados, como políticas, funções e responsabilidades, características de dados, financiamento à gestão de dados, documentação, compartilhamento e preservação de dados.

Dentre os resultados obtidos no estudo de Whitmire, Boock e Sutton (2015) relacionados à coleta de dados, destaca-se que em sete das doze faculdades investigadas, mais de 80% dos docentes mencionaram que estão produzindo dados em formato de imagem digital. Os autores apontaram que o predomínio de dados produzidos nesse formato, observado no *campus* da universidade, foi relatado em estudos anteriores (STEINHART *et al.*, 2012; AVERKAMP *et al.*, 2014). Tal predomínio tornou-se um fenômeno relevante neste contexto, na medida em que impacta a gestão de dados em processos relacionados ao armazenamento e *backup* de dados, à nomenclatura de arquivos, à atribuição de metadados e à preservação de dados.

Considerando o fato de que grandes quantidades de imagens digitais estão sendo criadas na universidade, emerge a necessidade potencial do fornecimento de materiais e de suporte especializados relacionados às melhores práticas para a gestão de imagens digitais, e um envolvimento mais evidente de docentes, no que se refere ao tópico da utilização de imagens digitais em suas pesquisas (WHITMIRE; BOOCK; SUTTON, 2015).

4.1.5.3 Processamento técnico de dados

Dentre os artigos analisados no estudo, cinco trataram a respeito da prática do processamento técnico de dados de pesquisa, os quais são apresentados no Quadro 14, e, posteriormente, serão abordados ao longo da narrativa.

Quadro 14 – Artigos que abordaram a prática do processamento técnico de dados de pesquisa

Prática	Título	Autor(es)	Ano
	ERDMAS: An exemplar-driven institutional research data management and analysis strategy	BELLGARD	2020
	Three camps, one destination: the intersections of research data management, FAIR and Open	HIGMAN; BANGERT; JONES	2019
Processar	Developing research data management services and support for researchers: a mixed methods study	PERRIER; BARNES	2018
	Variability in academic research data management practices: Implications for data services development from a faculty survey	WHITMIRE; BOOCK; SUTTON	2015
	Assessing Research Data Management Practices of Faculty at Carnegie Mellon University	VAN TUYL, MICHALEK	2015

Fonte: elaborado pela autora

Bellgard (2020) apresentou um modelo conceitual de gestão de dados de pesquisa baseado em exemplos e estratégia de análises, a partir de dois eixos. No primeiro, foram evidenciadas abordagens relativas a políticas de gestão de dados de pesquisa, diretrizes e procedimentos de auditoria. Já no segundo eixo, o foco foi direcionado ao pesquisador e às suas atividades de pesquisa.

Dentre os aspectos relacionados a práticas de gestão de dados de pesquisa mencionados pelo autor, destaca-se a relevância do processamento de dados, a partir de exemplo citado sobre dados obtidos em instrumentos de base hospitalar, tal como o tomógrafo. Segundo o autor, nesse contexto é necessário que os dados sejam submetidos a um meticuloso trabalho de anonimização em vários estágios da pesquisa, os quais exigem intervenção manual. Ademais, é importante que se estabeleçam os níveis de acesso aos dados e que se defina sua custódia, principalmente, no que se refere a posterior utilização em outros projetos de pesquisa (BELLGARD, 2020).

Em estudo que explorou as relações e os limites dos conceitos de “gestão de dados de pesquisa”, “princípio do FAIR” e “dados abertos”, Higman, Bangert e Jones (2019) apresentaram as seguintes definições para cada um desses conceitos:

- **gestão de dados de pesquisa** – conjunto de práticas utilizadas para lidar com dados coletados durante a pesquisa científica, tornando-os mais fáceis de serem localizados e compreendidos;

- **princípios FAIR** – formulados para facilitar a localização, acessibilidade, interoperabilidade e reutilização de dados de pesquisa;
- **dados abertos** – refere-se à disponibilização pública de dados de pesquisa, com restrições mínimas.

Considerando tais definições, Higman, Bangert e Jones (2019) destacaram a prática de processamento de dados como uma importante atividade para a obtenção de dados facilmente compartilháveis, mediante a disponibilização de metadados. Isso, principalmente, em casos nos quais conjuntos de dados encontram-se fechados e inacessíveis de outra forma, a não ser por meio de seus metadados.

Perrier e Barnes (2018) conduziram pesquisa para identificar as ferramentas essenciais e os serviços requeridos por pesquisadores para que a gestão de dados de pesquisa cumpra determinações estabelecidas por agências de financiamento e editores de periódicos científicos. Para isso, quatro grupos focais orientados sob paradigma pós-positivista foram realizados com 28 participantes, incluindo membros recrutados do corpo docente da Universidade de Toronto, coordenadores de pesquisa, estudantes de pós-doutorado, bibliotecários liderando pesquisas e profissionais de tecnologia que auxiliavam pesquisadores na gestão de seus dados. Esses participantes também classificaram serviços e ferramentas potenciais por meio de uma escala *Likert*, com a finalidade de aferir o grau de priorização entre os pesquisadores sobre tais aspectos. Os dados coletados por intermédio dos grupos focais foram codificados e, posteriormente, analisados.

Dentre os resultados obtidos no estudo de Perrier e Barnes (2018), salienta-se a preocupação dos pesquisadores em relação ao adequado processamento de dados que tratam a respeito de seres humanos. Sobre isso, práticas como a anonimização de dados são extremamente relevantes, conforme apontado por um participante da pesquisa:

Tínhamos muitas preocupações em termos de privacidade e confidencialidade e em tentar desidentificar as pessoas (...) se quisermos ter todos os dados disponíveis, esses tipos de problemas serão muito, muito grandes só porque são informações pessoais e identificáveis (GRUPO DE FOCO 1, PARTICIPANTE 1, 2018 *apud* PERRIER; BARNES, 2018, *tradução nossa*).

Quanto ao processamento de dados de pesquisa, especialmente no que se refere à atribuição de metadados, Whitmire, Boock e Sutton (2015) constataram que docentes atuantes em diversas áreas da *Oregon State University* (OSU) têm realizado essa prática. Dentre os resultados obtidos no estudo dos autores, 88% dos docentes na área de Ciências da Terra, Oceânica e Atmosférica relataram a criação de metadados para os dados que produzem. Todavia, apenas 19% dos docentes investigados na área da Medicina Veterinária afirmaram atribuir metadados a seus dados.

Assim também, docentes que relataram o uso de metadados foram questionados sobre os padrões que estão adotando. Sobre esse aspecto, Whitmire, Boock e Sutton (2015) observaram a utilização de padrões como: DC (*Dublin Core*); DwC (*Darwin Core*); DDI (*Data Documentation Initiative*); DIF (*Directory Interchange Format*); EML (*Ecological Metadata Language*); FGDC (*Federal Geographic Data Committee*) e ISO 19115. Entretanto, verificou-se que muitos docentes estão criando metadados conforme esquemas estabelecidos em seus grupos de pesquisa. Para os referidos autores, conjuntos de dados com metadados padronizados são mais interoperáveis com outros dados, e mais legíveis por máquinas e humanos.

De modo semelhante, na pesquisa realizada por Van Tuyl e Michalek (2015) no contexto de pesquisadores da *Carnegie Mellon University* (CMU), os autores observaram diversos métodos relacionados ao processamento de dados de pesquisa. Nesse sentido, entrevistados no estudo relataram métodos utilizados para a documentação de dados e inclusão de metadados, o que abrangia tanto arranjos informais, quanto arranjos fortemente organizados com a adoção de documentação formal, padrões de metadados e regras compartilhadas em relação a nomes e registros de arquivos.

Outros aspectos abordados entre os entrevistados sobre a inclusão de metadados e documentação foram destacados por Van Tuyl e Michalek (2015):

- pesquisadores ainda não detinham conhecimentos suficientes para a criação de metadados;
- havia dados considerados “auto-descritivos” e não exigiam atribuição de metadados;

- alguns dados não seriam reutilizados, ainda que fossem bem documentados.

4.1.5.4 Compartilhamento de dados

Com relação à prática de compartilhamento de dados de pesquisa e a gestão dos mesmos, três artigos analisados no estudo discutiram esse tópico específico, conforme apresentado no Quadro 15. Logo em seguida, tais artigos serão abordados nesta seção.

Quadro 15 – Artigos que abordaram a prática do compartilhamento de dados de pesquisa

Prática	Título	Autor(es)	Ano
Compartilhar	FAIR x FAIR. Una aproximación pragmática a la gestión de datos de investigación en modo FAIR	ALCALÁ; ANGLADA	2019
	Research data management practices of faculty members	PIRACHA; AMEEN	2018
	Variability in academic research data management practices: Implications for data services development from a faculty survey	WHITMIRE; BOOCK; SUTTON	2015

Fonte: elaborado pela autora

Por meio de análise documental e de entrevistas realizadas com docentes de instituições de ensino em Catalunha, Alcalá e Anglada (2019) salientaram duas alternativas já apontadas pelo *Consorci de Serveis Universitaris de Catalunya* (CSUC) como maneiras de compartilhar dados abertos de pesquisa. A primeira referiu-se à publicação de dados em repositórios destinados exclusivamente a dados, enquanto a segunda, se tratava da adaptação de repositórios institucionais já existentes para o depósito de dados produzidos nas universidades. Entretanto, essa última não cumpriria com todos os princípios FAIR, devido a aspectos como ausência de indicadores, preservação limitada ao contexto de dados e baixa capacidade de armazenamento.

Tomando como pano de fundo o exposto, Alcalá e Anglada (2019) também discutiram acerca dos requisitos razoáveis os quais deveriam ser cumpridos por um repositório de dados, principalmente com base em entrevistas realizadas com 32 especialistas. Nesse sentido, foram destacados os seguintes aspectos: torna-se necessária a atribuição e utilização de identificadores persistentes, como DOIs para

dados e Orcid para pesquisadores; o repositório de dados deve suportar até 10 GB por padrão, podendo expandir sua capacidade de armazenamento; arquivos contendo dados devem estar disponíveis por 10 anos, sendo geograficamente distribuídas duas cópias desses arquivos; repositórios devem se comunicar com outros repositórios e instrumentos de armazenamento em nuvem, de modo a possibilitar um carregamento amigável de dados.

Piracha e Ameen (2018) conduziram estudo cujo objetivo foi investigar práticas de gestão de dados de pesquisa realizadas no âmbito da Universidade do Punjab. Para tanto, dados qualitativos foram obtidos por meio de entrevistas semiestruturadas com dez docentes da universidade. A amostra foi intencionalmente selecionada com base em registros de publicação de pesquisas e incluiu docentes atuantes nas áreas de Ciências da Terra, Bioquímica, Microbiologia, Engenharia Química, Matemática, Ciências Administrativas, Botânica, Zoologia e Medicina.

Dentre os resultados alcançados no estudo, verificou-se que todos os entrevistados reagiram positivamente à prática de compartilhamento de dados, e afirmaram estar de acordo com a disponibilização de seus dados em repositórios. Também apreciaram a sugestão da instituição de um repositório de dados gerido pela biblioteca central da universidade. Um dos entrevistados chegou a afirmar que se tal medida tivesse sido implementada antes, ele não teria perdido seus dados de pesquisa obtidos em 50 anos de trabalho (PIRACHA; AMEEN, 2018).

Considerando o crescimento na prática do compartilhamento de dados realizado por pesquisadores acadêmicos em repositórios institucionais, Whitmire, Boock e Sutton (2015) destacaram o papel de bibliotecas da *Oregon State University* (OSU) que estão implementando uma nova plataforma para o repositório institucional da universidade. Tal medida implicará em metadados mais expansivos e explicitará relacionamentos entre os objetos que irão compor o repositório. Nessa direção, pretende-se estimular o depósito de dados por meio da adoção de melhores práticas de gestão de dados, de modo a assegurar a preservação de arquivos.

4.1.5.5 Preservação de dados

No que diz respeito à prática de preservação de dados de pesquisa, três artigos investigados no estudo abordaram esse tópico específico, como pode ser visto no Quadro 16. Logo após, tais artigos serão abordados na seção.

Quadro 16 – Artigos que abordaram a prática da preservação de dados de pesquisa

Prática	Título	Autor(es)	Ano
Preservar	Research data management at a public university in Malawi: the role of “three hands”	CHAWINGA; ZINN	2020
	Developing research data management services and support for researchers: a mixed methods study	PERRIER; BARNES	2018
	Variability in academic research data management practices: Implications for data services development from a faculty survey	WHITMIRE; BOOCK; SUTTON	2015

Fonte: elaborado pela autora

Chawinga e Zinn (2020) conduziram pesquisa cujo objetivo foi investigar práticas de gestão de dados de pesquisa em uma universidade pública de Malawi (nome da universidade omitido por razões éticas) para propor estratégias de gestão adequadas. Para isso, os dados do estudo foram coletados por meio de questionário respondido por integrantes da equipe da biblioteca e por pesquisadores, bem como por entrevistas realizadas com o diretor de pesquisa da universidade.

Resultados do estudo de Chawinga e Zinn (2020) demonstraram o quão relevante é a prática da preservação de dados em direção a uma gestão de dados adequada, uma vez que se observou na universidade mencionada a carência de infraestruturas para esse fim. Com efeito, pesquisadores armazenavam os dados que produziam utilizando ferramentas básicas, o que incluía computadores pessoais, discos rígidos externos e contas de correio eletrônico. Dentre os aspectos mais relevantes em relação a isso, ficou exposta a fragilidade e vulnerabilidade de dados de pesquisadores devido a danos acidentais, extravio ou furtos, e a ausência de repositórios de dados institucionais.

Semelhantemente, estudo realizado por Perrier e Barnes (2018) na Universidade de Toronto, apontou que pesquisadores armazenavam seus dados conforme diversas soluções individuais que abrangiam banco de dados não proprietários. Além disso, pesquisadores expressaram o desejo de que a universidade dispusesse de infraestrutura adequada ao armazenamento de dados, e que a preservação desses fosse tópico abordado em nível institucional.

Déficits, no que se refere ao armazenamento de dados, também foram identificados por Whitmire, Boock e Sutton (2015), no âmbito da *Oregon State University* (OSU). Na investigação dos autores, docentes da OSU relataram que na maioria das vezes armazenavam dados em computadores pessoais e dispositivos de armazenamento externo. Além disso, tomou-se nota de que em vários departamentos e faculdades, docentes informaram que armazenavam dados em servidores mantidos pelo grupo de pesquisa os quais eram membros, apesar de sua faculdade ou departamento dispor de servidor em rede destinado a esse fim. As Faculdades de Engenharia, Ciências e Medicina Veterinária se destacaram por expressivo número de docentes que utilizavam servidores do próprio grupo de pesquisa para o armazenamento de dados (em todas essas faculdades o número correspondeu a mais de 50% dos docentes inquiridos).

4.1.5.6 Descoberta de dados de pesquisa

Dentre os artigos analisados no estudo, dois discutiram aspectos relacionados ao aumento da visibilidade de dados de pesquisa proporcionado por efetiva gestão dos mesmos, fazendo-os serem descobertos por outros pesquisadores. O Quadro 17 descreve quais são esses artigos que, posteriormente, serão explorados ao longo da narrativa.

Quadro 17 – Artigos que abordaram a descoberta de dados de pesquisa

Prática	Título	Autor(es)	Ano
Descobrir	Three camps, one destination: the intersections of research data management, FAIR and Open	HIGMAN; BANGERT; JONES	2019
	Assessing Research Data Management Practices of Faculty at Carnegie Mellon University	VAN TUYL, MICHALEK	2015

Fonte: elaborado pela autora

De acordo com Higman, Bangert e Jones (2019), a inclusão de metadados em conjuntos de dados contribui para torná-los mais visíveis e localizáveis por outros pesquisadores que os descobrirão. Consequentemente, dados podem ser reutilizados ativamente em contextos distintos dos quais foram criados.

No que se refere à visibilidade e descoberta de conjuntos de dados de pesquisa, Van Tuyl e Michalek (2015) relataram algumas ações implementadas no contexto da *Carnegie Mellon University* (CMU). Dentre essas, pode-se destacar o desenvolvimento de um repositório de dados na universidade, visando promover a gestão, preservação, descoberta e acesso a dados, e o apoio à criação de DOI's para conjuntos de dados, vinculando-os a todas as produções acadêmicas de pesquisadores ao longo do tempo.

4.1.5.7 Reuso de dados de pesquisa

Três dos artigos analisados no estudo abordaram a prática do reuso de dados de pesquisa associada à gestão dos mesmos, os quais são apresentados no Quadro 18, e, posteriormente, serão explorados ao longo da narrativa.

Quadro 18 – Artigos que abordaram a prática do reuso de dados de pesquisa

Prática	Título	Autor(es)	Ano
Reuso	Research data management at a public university in Malawi: the role of “three hands”	CHAWINGA; ZINN	2020
	Three camps, one destination: the intersections of research data management, FAIR and Open	HIGMAN; BANGERT; JONES	2019
	FAIR x FAIR. Una aproximación pragmática a la gestión de datos de investigación en modo FAIR	ALCALÁ; ANGLADA	2019

Fonte: elaborado pela autora

Em pesquisa conduzida por Chawinga e Zinn (2020) no âmbito de uma universidade pública de Malawi, foi observado dentre os resultados alcançados, que o reuso de dados não é frequentemente realizado por pesquisadores. Assim também, três desafios foram elencados em relação a essa prática: dificuldades no acesso a dados reutilizáveis; falta de metadados padronizados; receio de que dados possam ser mal interpretados. A partir do exposto, é possível inferir que medidas associadas à facilitação no acesso a dados, à atribuição de metadados e à realização de análises adequadas a respeito de dados contribuem para a ampliação da prática do reuso desses insumos entre pesquisadores.

Explorando a relação dos conceitos de “gestão de dados de pesquisa” “princípio do FAIR” e “dados abertos”, Higman, Bangert e Jones (2019) afirmaram que o valor correspondente aos dados é maximizado quando a abertura e os princípios FAIR são atingidos em nível elevado. Todavia, apesar de abertos e estarem em conformidade com tais princípios, se não gerenciados adequadamente, resultam em dados mal documentados e potencialmente menos reutilizáveis. Outro aspecto igualmente relevante nesse contexto, é a identificação de maneira exclusiva de dados abertos de pesquisa, o que os torna mais visíveis e contribui para o aumento na prática de seu reuso.

De acordo com Alcalá e Anglada (2019), não há dúvidas de que dispor de infraestrutura apropriada em um repositório de dados é conveniente para promover o acesso e reutilização de dados. Entretanto, isso não é suficiente, uma vez que a adoção de práticas eficazes de gestão de dados de pesquisa é importante nesse contexto e inclui: a organização de dados de modo a torná-los compreensíveis e

reutilizáveis por usuários; o encorajamento do uso de formatos abertos, sempre que possível; a utilização de padrões e vocabulários controlados comumente aceitos.

4.1.5.8 Monitoramento e controle de dados de pesquisa

Com relação à prática de monitoramento e controle de dados de pesquisa, apenas um artigo foi identificado acerca desse tópico específico, como apresentado no Quadro 19.

Quadro 19 – Artigo que abordou a prática do monitoramento e controle de dados de pesquisa

Prática	Título	Autor(es)	Ano
Monitoramento e controle	Assessing Research Data Management Practices of Faculty at Carnegie Mellon University	VAN TUYL; MICHALEK	2015

Fonte: elaborado pela autora

Quanto à prática de monitoramento e controle relacionada à gestão de dados de pesquisa, Van Tuyl e Michalek (2015) a ilustraram a partir de iniciativa realizada em 2011, pelas bibliotecas da *Carnegie Mellon University (CMU)*, juntamente com serviços de computação e membros do escritório de pesquisa da universidade. Nessa ocasião, um grupo de trabalho foi instaurado, sendo responsável por conduzir avaliações informais regulares sobre o cenário referente aos dados produzidos na CMU, e por acompanhar o desenvolvimento de iniciativas de dados em instituições no mundo. Tais avaliações foram resultados de discussões com professores e profissionais de tecnologia na universidade, de um estudo piloto realizado com docentes a fim de analisar práticas de gestão de dados, e de uma investigação ambiental sobre como outras instituições estavam iniciando a prestação de serviços nessa área.

4.1.5.9 Elaboração de políticas para a gestão de dados de pesquisa

Quatro artigos analisados no estudo mencionaram a prática da elaboração de políticas para a gestão de dados de pesquisa. Esses artigos são apresentados no Quadro 20, e, logo em seguida, serão detalhadamente abordados.

Quadro 20 – Artigos que abordaram a prática da elaboração de políticas para a gestão de dados de pesquisa

Prática	Título	Autor(es)	Ano
Políticas de gestão de dados de pesquisa	Research Data Management (RDM) Capabilities at the University of Ghana, Legon	AVUGLAH; UNDERWOOD	2019
	Cultural obstacles to research data management and sharing at TU Delft	PLOMP et al.	2019
	Research data management in Turkey: perceptions and practices	AYDINOGLU; DOGAN; TASKIN	2017
	Research data management and openness: The role of data sharing in developing institutional policies and practices	HIGMAN; PINFIELD	2015

Fonte: elaborado pela autora

Segundo Avuglah e Underwood (2019), a participação de governos e organizações de financiamento à pesquisa é o principal elemento de contribuição para a gestão de dados de pesquisa no âmbito de instituições de ensino superior. Entretanto, via de regra, isso não ocorre no continente africano. A partir desse pano de fundo, os autores conduziram pesquisa com a finalidade de avaliar as capacidades de gestão de dados de pesquisa na Universidade de Gana (UG). Para a coleta de dados, o estudo lançou mão de entrevistas semiestruturadas realizadas com integrantes do Sistema de Bibliotecas da Universidade de Gana, do Sistema de Computação da universidade, dois membros do Escritório de Pesquisa, Inovação e Desenvolvimento da instituição e dois pesquisadores seniores. Os dados obtidos foram analisados conforme codificação das entrevistas e também foram realizadas análises de documentos.

Dentre os resultados obtidos no estudo de Avuglah e Underwood (2019), destaca-se a ausência de uma política formal relativa à gestão de dados de pesquisa na universidade. Apesar disso, os autores identificaram um documento institucional

que abrangeu um conjunto de melhores práticas orientadas à gestão de dados, intitulado: “Diretriz de Política de Pesquisa da (UG) sobre boas práticas: manutenção de registros e gestão de dados”. Tal documento apresentou aspectos relevantes sobre esse tópico e incluiu informações acerca da propriedade de dados, produção, documentação, armazenamento e compartilhamento de dados.

A diretriz mencionada também abordou a política do repositório de dados da instituição. Dentre os aspectos tratados, pode-se salientar a adoção de metadados para a descrição de conjuntos de dados, orientações acerca de serviços institucionais, padrões de operação e suporte a itens depositados, assim como instruções relacionadas à propriedade intelectual de dados (AVUGLAH; UNDERWOOD, 2019).

No que se refere à elaboração de políticas de dados de pesquisa, Plompe *et al.* (2019) mencionaram a publicação da Política de Estrutura de Dados de Pesquisa da *Delft University of Technology* (TU Delft), em 2018. Essa política descreveu as atribuições relativas à gestão de dados incubidas à biblioteca, ao departamento de tecnologia da informação e comunicação, à Escola de Pós-Graduação e ao Conselho Executivo da universidade. Assim também, tal documento considerou aspectos relativos às diferenças disciplinares.

Aydinoglu, Dogan e Taskin (2017) conduziram pesquisa com o propósito de investigar as percepções e práticas de pesquisadores turcos em relação à gestão de dados de pesquisa. Os dados do estudo foram obtidos por meio da aplicação de questionário a 532 acadêmicos, provenientes das 25 universidades turcas mais produtivas. A análise dos dados ocorreu com o auxílio do pacote de *software IBM SPSS Statistics* (v. 21). As áreas nas quais os participantes da pesquisa atuavam incluíam Ciências Humanas e Sociais, Engenharia, Ciências da Saúde, Agricultura e pesca e Ciências.

A investigação de Aydinoglu, Dogan e Taskin (2017) verificou a ausência de políticas de gestão de dados em nível institucional. A principal agência de financiamento à pesquisa na Turquia – denominada TUBITAK – não dispõe desse instrumento, que, por sua vez, também não é comum no contexto de universidades. Entretanto, como explicaram os autores, tais políticas são importantes principalmente para que pesquisadores turcos se orientem e adotem as melhores práticas de gestão

dos dados em relação aos dados que produzem, e estejam em consonância com exigências estabelecidas por programas internacionais de financiamento à pesquisa.

Em estudo realizado por Higman e Pinfield (2015) acerca de políticas e práticas de gestão de dados de pesquisa em Instituições de Ensino Superior (IES) do Reino Unido, foi verificado que muitas políticas analisadas não detalhavam, suficientemente, procedimentos relativos à gestão de dados, ou promoviam uma vaga menção em relação a esse tópico. Por outro lado, observou-se que a produção de políticas foi significativamente influenciada por requisitos estabelecidos por financiadores da pesquisa, sendo objeto de discussão apontado por pesquisadores.

4.1.5.10 Suporte à análise de dados

Com relação à prática de suporte à análise de dados de pesquisa, somente um artigo foi identificado acerca desse tópico específico, como apresentado no Quadro 21.

Quadro 21 – Artigo que abordou a prática de suporte à análise de dados de pesquisa

Prática	Título	Autor(es)	Ano
Suporte à análise de dados	Research Data Management (RDM): Capabilities at the University of Ghana, Legon	AVUGLAH; UNDERWOOD	2019

Fonte: elaborado pela autora

Uma ilustração acerca de suporte institucional a pesquisadores com relação à análise de dados é mencionada por Avuglah e Underwood (2019), no contexto da Universidade de Gana (UG). Assim, os autores observaram que a universidade oferecia apoio a pesquisadores no que se refere à análise de dados, por meio de treinamento na utilização de *softwares* que auxiliam essa função, tais como SPSS, NVivo e similares.

4.1.5.11 Treinamento sobre a gestão de dados de pesquisa orientado a profissionais de informação

Dentre os artigos analisados no estudo, dois abordaram a prática da realização de treinamento sobre a gestão de dados de pesquisa orientado a profissionais de informação. Tais artigos são apresentados no Quadro 22, e, posteriormente, serão explorados ao longo da narrativa.

Quadro 22 – Artigos que abordaram a prática da realização de treinamento sobre a gestão de dados de pesquisa orientado a profissionais de informação

Prática	Título	Autor(es)	Ano
Treinamento a profissionais de informação	The collaborative creation of a Research Data Management Librarian Academy (RDMLA)	SHIPMAN; TANG	2019
	Moving a brick building: UK libraries coping with research data management as a 'wicked' problem	COX; PINFIELD; SMITH	2016

Fonte: elaborado pela autora

No que se refere à realização de treinamentos *online* gratuitos direcionados a profissionais de informação sobre melhores práticas de gestão de dados de pesquisa, Shipman e Tang (2019) destacaram capacitação promovida pela *Research Data Management Librarian Academy* (RDMLA), resultado de uma parceria entre a empresa *Elsevier* e equipe composta por bibliotecários da área de ciências da saúde, administradores de bibliotecas e membros do corpo docente em *Library and Information Science* (LIS) de Boston.

A equipe responsável pelo treinamento avaliou necessidades relacionadas a esse tópico e identificou lacunas, no que tange a oportunidades de treinamento. Dessa forma, elaborou currículo composto por oito unidades para o treinamento, as quais incluem: 1) fundamentos da gestão de dados de pesquisa; 2) a cultura de dados de pesquisa; 3) defesa e divulgação do valor da gestão de dados de pesquisa em bibliotecas; 4) lançamento de serviços de dados em bibliotecas; 5) gestão e avaliação de projetos; 6) visão geral da análise de dados e ferramentas de visualização; 7) visão geral das ferramentas de codificação; 8) visão geral de ferramentas de plataformas.

Cox, Pinfield e Smith (2016) retomaram o conceito de “problema perverso” proposto por Rittel e Webber (1973) para explorar temática referente à gestão de dados de pesquisa. Nessa perspectiva, um problema perverso pode ser entendido como um desafio único, demasiadamente complexo, cuja solução se manterá desconhecida.

Com base nisso, Cox, Pinfield e Smith (2016) conduziram pesquisa com o objetivo de investigar o valor para os bibliotecários em entender a gestão de dados de pesquisa como um problema "perverso". Para tanto, o estudo lançou mão de 26 entrevistas semiestruturadas com bibliotecários a respeito de serviços de gestão de dados de pesquisa em bibliotecas universitárias do Reino Unido. As gravações das entrevistas foram transcritas, e em seguida codificadas com o auxílio do *software* NVivo. Dentre os resultados obtidos, destaca-se que serviços de treinamento e consultoria orientados a bibliotecários foram considerados fundamentais para uma gestão de dados de pesquisa adequada, uma vez que se trata de um problema complexo que acarreta mudanças culturais significativas no contexto dessa profissão (COX; PINFIELD; SMITH, 2016).

4.1.5.12 Treinamento sobre a gestão de dados de pesquisa orientado a pesquisadores

De modo semelhante, três artigos analisados no estudo abordaram a prática da realização de treinamento sobre a gestão de dados de pesquisa orientado a pesquisadores. Esses artigos são apresentados no Quadro 23, e, logo em seguida, serão detalhados.

Quadro 23 – Artigos que abordaram a prática da realização de treinamento sobre a gestão de dados de pesquisa orientado a pesquisadores

Prática	Título	Autor(es)	Ano
Treinamento a pesquisadores	Research data management practices of faculty members	PIRACHA; AMEEN	2018
	Research data management in Turkey: perceptions and practices	AYDINOGLU; DOGAN; TASKIN	2017
	Assessing Research Data Management Practices of Faculty at Carnegie Mellon University	VAN TUYL, MICHALEK	2015

Fonte: elaborado pela autora

Ao analisarem práticas de gestão de dados de pesquisa realizadas no âmbito da Universidade do Punjab, Piracha e Ameen (2018) observaram que docentes entrevistados no estudo relataram que a gestão de dados se tornou um trabalho técnico, sendo fundamental o domínio de habilidades que assegurem o acesso futuro a dados. Além disso, apontaram a necessidade de obtenção de apoio da equipe da biblioteca para assimilarem novas técnicas relacionadas à curadoria de dados.

Enfatizando a realização de treinamentos institucionais a pesquisadores turcos como um recurso relevante para a gestão de dados de pesquisa, Aydinoglu, Dogan e Taskin (2017) recomendaram que cursos de ciência de dados sejam integrados ao currículo dos departamentos de universidades, e que sejam organizados seminários e *workshops* extracurriculares sobre a temática.

Nessa direção, Van Tuyl e Michalek (2015) mencionaram que, em 2012, foram desenvolvidos serviços-piloto direcionados a pesquisadores e estudantes da *Carnegie Mellon University* (CMU) que incluíam *workshops* sobre gestão de dados, e discussões com administradores de pesquisa e professores acerca do compartilhamento de informações e de dados em repositórios. Isso contribui como insumo para a criação futura de serviços de dados na universidade, na medida em que ampliou a compreensão sobre a profundidade das necessidades relativas aos dados na instituição. Além disso, no que se refere à gestão de dados de pesquisa, os autores salientaram que é preciso adotar uma abordagem diferenciada e multidisciplinar para a capacitação de pesquisadores.

Considerando o exposto, o Quadro 24 apresenta a síntese das práticas de gestão de dados de pesquisa em diferentes áreas, as quais foram identificadas na pesquisa. Nele, pode-se observar uma breve descrição de cada prática e a referência dos trabalhos em que foram localizadas:

Quadro 24 – Síntese das práticas de gestão de dados de pesquisa em diferentes áreas

Prática	Descrição	Trabalhos identificados
Planejamento	Atividades de planejamento da gestão de dados e de resultados de pesquisas	PLOMP <i>et al.</i> , 2019; ALCALÁ; ANGLADA, 2019; HIGMAN; PINFIELD, 2015; VAN TUYL; MICHALEK, 2015
Coleta	Procedimentos relacionados à produção/obtenção de dados	WHITMIRE; BOOCK; SUTTON, 2015
Processamento técnico	Atividades relacionadas ao tratamento técnico de dados de pesquisa, como a inclusão de metadados, anonimização, digitalização e tradução de dados	BELLEGARD, 2020; HIGMAN; BANGERT; JONES, 2019; PERRIER; BARNES, 2018; WHITMIRE; BOOCK; SUTTON, 2015; VAN TUYL; MICHALEK, 2015
Compartilhamento	Procedimentos e sistemas que viabilizam o adequado compartilhamento de dados de pesquisa	ALCALÁ; ANGLADA, 2019; PIRACHA; AMEEN, 2018; WHITMIRE; BOOCK; SUTTON, 2015
Preservação	Procedimentos e sistemas que proporcionam a adequada preservação de dados de pesquisa a longo prazo	CHAWINGA; ZINN, 2020; PERRIER; BARNES, 2018; WHITMIRE; BOOCK; SUTTON, 2015
Descoberta	Adoção de estratégias que contribuem para o aumento da visibilidade de dados de pesquisa	HIGMAN; BANGERT; JONES, 2019; VAN TUYL; MICHALEK, 2015
Reuso	Procedimentos que maximizam o potencial reuso de dados de pesquisa	CHAWINGA; ZINN, 2020; HIGMAN; BANGERT; JONES, 2019; ALCALÁ; ANGLADA, 2019 VAN TUYL; MICHALEK, 2015
Monitoramento e controle	Inclui a avaliação da gestão de dados e o monitoramento de atividades realizadas ao longo da pesquisa científica	
Políticas para a gestão de dados de pesquisa	Elaboração de documentos formais que orientem quais procedimentos devem ser adotados em relação à gestão de dados de pesquisa	AVUGLAH; UNDERWOOD, 2019; PLOMP <i>et al.</i> 2019; AYDINOGLU; DOGAN; TASKIN, 2017; HIGMAN; PINFIELD, 2015
Suporte à análise de dados	Refere-se ao apoio institucional destinado a pesquisadores em relação a análise de dados	AVUGLAH; UNDERWOOD, 2019
Treinamento oferecido a profissionais de informação	Treinamento sobre a gestão de dados de pesquisa orientado a profissionais de informação	SHIPMAN; TANG, 2019; COX; PINFIELD; SMITH, 2016
Treinamento oferecido a pesquisadores	Treinamento sobre a gestão de dados de pesquisa orientado a pesquisadores	PIRACHA; AMEEN, 2018; AYDINOGLU; DOGAN; TASKIN, 2017; VAN TUYL; MICHALEK, 2015

Fonte: elaborado pela autora

Após a caracterização dos artigos investigados na pesquisa e do levantamento de práticas de gestão de dados de pesquisa em diferentes áreas, com base na literatura científica analisada, a próxima seção continuará abordando tais práticas, porém, sob a perspectiva de pesquisadores em diferentes áreas.

4.2 Analisar práticas de gestão de dados de pesquisa realizadas por pesquisadores em diferentes áreas

As práticas de gestão de dados de pesquisa realizadas por pesquisadores na área de Sociologia e de Química serão apresentadas conforme os resultados obtidos em cada um dos processos inerentes à teoria fundamentada, e, especificamente, nessa sequência: 1) codificação aberta; 2) codificação axial; 3) codificação seletiva e teoria substantiva. Ressalta-se, ainda, que a primeira área a ser abordada é a Sociologia, e logo em seguida, a da Química.

4.2.1 Codificação aberta - Sociologia

Esta seção apresentará a síntese dos principais aspectos abordados pelos pesquisadores na área de Sociologia que foram entrevistados no estudo. Além disso, descreverá as dimensões, categorias e propriedades identificadas por meio de processo analítico denominado: codificação aberta. Toma-se nota, ainda, que Strauss e Corbin (2008) definem esse processo como a identificação e nomeação de conceitos contidos em um texto, os quais são capazes de representar fenômenos. Tais conceitos foram chamados de categorias.

Com o propósito de anonimizar os pesquisadores entrevistados, todas as vezes que esses forem citados no texto, serão mencionados com o código “PSnº”, em que “P” refere-se ao termo “pesquisador”, “S” ao termo “Sociologia” e “nº” ao número sequencial atribuído a cada pesquisador, conforme a ordem de realização das entrevistas.

4.2.1.1 Codificação aberta da entrevista com PS1

PS1 relatou que o tópico “Ciência aberta” ainda não tem sido amplamente discutido, no sentido de estar sendo pautado em eventos científicos de grande visibilidade na área. Logo, essa temática tem sido abordada, principalmente, de modo informal entre pesquisadores.

Com relação ao planejamento das pesquisas que realiza, PS1 disse que procura otimizar recursos e reduzir custos na realização de seus estudos. Para tanto, estima recursos e solicita financiamento a agências de fomento à pesquisa e à universidade a qual está vinculado. Todavia, revelou a dificuldade de propor pesquisas originais com dados específicos, considerando a falta de estímulos ao pesquisador, que apesar de seu envolvimento com a produção científica, comumente não é reconhecido pelo conjunto de responsabilidades jurídicas, éticas e burocráticas as quais está submetido.

PS1 também afirmou que já elaborou um Plano de gestão de dados de pesquisa, o que lhe gerou muitas dificuldades, em razão de não dispor de algum apoio que pudesse lhe orientar nesse sentido. No que se refere à documentação de processos relativos aos dados que produz, PS1 disse que tal medida é fundamental para a preservação do próprio histórico da pesquisa, principalmente, devido à rotatividade de pessoas no ambiente de produção científica. Entretanto, no que diz respeito à preservação de dados, essa ainda é realizada por meio de estratégias locais, o que inclui o armazenamento de dados em computadores, servidores locais, *Hard Disk* (HD) e mídias como *Digital Versatile Disc* (DVD) e *Compact Disc* (CD).

Ao comentar sobre o compartilhamento de dados de pesquisa, PS1 destacou que essa é uma prática que geralmente ocorre no escopo do grupo de pesquisa que participa, no qual há colaboração científica. Os trabalhos resultantes dessa colaboração costumam ser publicados em coautoria, como se pode notar:

No grupo, a pessoa, em geral, te pede acesso aos dados e te convida para analisar juntos. Então, e isso é bom para a pessoa, eu te peço acesso aos dados e eu te convido para analisá-los junto comigo. Aí, eu entro lá como sexto autor, por exemplo [...] e aí eu assumo a responsabilidade de resolver os problemas com os dados. Então, a minha colaboração com a pesquisa é fornecer os dados e resolver dilemas com os dados.

PS1 ainda relatou que já teve seus dados de pesquisa disponibilizados por instituições que fomentaram determinado estudo, porém, não o fez em nível individual, principalmente, devido à insuficiência de recursos tecnológicos e financeiros que viabilizassem o compartilhamento de dados. Por sua vez, o reuso de dados de pesquisa concentra-se no contexto de grupos nos quais atua em rede de pesquisa.

No que se refere à gestão de dados abertos em universidades, PS1 destacou que as instituições brasileiras não dispõem de acordos ou diretrizes que orientem como pesquisadores devem proceder nesse sentido. Além disso, há dilemas éticos,

jurídicos e burocráticos, que somados à insuficiência de recursos, dificultam tal prática, como é possível perceber:

Conseguir resolver os dilemas éticos, jurídicos e burocráticos, é muito difícil porque é muito caro. Você tem que ter um site, tem que ter um bolsista ou um profissional, um técnico para alimentar o site, você tem que ter gente para produzir documentação, gente para atender as pessoas que tem dúvida sobre aquilo. Eu acho muito difícil, pelo menos nas ciências humanas e sociais [...] eu acho que a universidade pode até querer que os professores, que seus funcionários assumam toda essa responsabilidade. Eu acho que na verdade, ela vai até querer isso, mas a universidade assumir essa responsabilidade, eu duvido.

Três dilemas foram apontados por PS1 em relação aos dados de pesquisa: 1) assegurar o respeito a acordos de confidencialidade de dados; 2) apropriação indevida de dados e 3) perda do ineditismo de trabalhos científicos. O primeiro refere-se a acordos realizados com participantes da pesquisa, nos quais nem sempre é permitida a divulgação de dados. O segundo, diz respeito ao receio de o pesquisador ter seus dados reutilizados por outros, sem a devida atribuição de autoria. E o último aspecto trata da perda do ineditismo na publicação de trabalhos científicos, na medida em que o acesso a dados de pesquisa poderia acirrar uma competição entre pesquisadores sobre quem publica artigos mais rapidamente.

Assim sendo, o Quadro 25 expõe as dimensões, categorias e propriedades identificadas no relato de PS1, por meio de codificação aberta:

Quadro 25 – Codificação aberta da entrevista com PS1: práticas de gestão de dados de pesquisa

Dimensão	Categoria	Propriedade
Discussão sobre Ciência aberta	Entre pesquisadores	Discussão informal, em menor escala
Aspectos éticos	Acordos de confidencialidade de dados	Anonimização de dados
	Indicação de responsabilidade de dados	Apropriação indevida de dados
	Publicação antecipada de dados	Perda do ineditismo
Planejamento	Custos	Estimativa de recursos
	Falta de estímulo ao pesquisador	Falta de reconhecimento
	Plano de gestão de dados	Elaboração de documento
Documentação de dados	Histórico de pesquisa	Processos da pesquisa
Preservação de dados	Estratégias locais	Realizada pelos próprios responsáveis da pesquisa
Compartilhamento de dados	Grupo de pesquisa	Ocorre no escopo do grupo de pesquisa
	Instituições de fomento à pesquisa	Solicitado por instituições de fomento à pesquisa
	Insuficiência de recursos	Meios que viabilizam o compartilhamento de dados
Reuso de dados	Grupos em rede de pesquisa	Colaboração científica
Gestão de dados abertos em universidades	Insuficiência de recursos	Meios que viabilizam a gestão de dados
	Falta de diretrizes	Orientações para a gestão de dados de pesquisa

Fonte: elaborado pela autora

4.2.1.2 Codificação aberta da entrevista com PS2

PS2 destacou que a discussão sobre o tópico “Ciência aberta” varia conforme o entendimento de cada área do conhecimento, e que essa é uma questão essencial.

A título de ilustração, ele disse:

Depende muito da área, eu acho que isso é uma questão central [...] tu vais para área das [ciências] médicas, das engenharias, da engenharia de materiais, a engenharia civil, a engenharia elétrica, aí, é um outro tipo de coisa, desenvolvem protótipo. Por exemplo, desenvolvem patente ou registro de patentes. Então, não é o meu caso, eu posso falar pela minha área, e é como eu me sinto bem.

No que se refere ao planejamento das pesquisas que conduz, PS2 afirmou que é rara a obtenção de financiamento à pesquisa por meio de iniciativas privadas, e que a concorrência nas participações de editais públicos é muito grande. Com indignação, PS2 evidenciou a falta de estímulo à produção científica, quando expôs suas dificuldades na obtenção de recursos:

Eu estou no auge da minha capacidade física, intelectual, e, de repente, eu tenho que me ver em situações assim, como se a estrutura, o sistema não me dá condições [...] eu vou continuar batalhando para conseguir os recursos para financiar as pesquisas, mas é um processo muito difícil [...] então, a gente vive dessa paixão e da vontade de ajudar de fazer alguma coisa, é muito difícil, entendeu?

PS2 ainda não teve experiência na elaboração de um Plano de gestão de dados de pesquisa. Quanto à coleta de dados, PS2 destacou a diversidade de tipos de dados que produz ou faz uso, os quais são obtidos por diferentes técnicas em contextos metodológicos qualitativos e quantitativos. Nessa direção, lança mão de fontes censitárias, tais como amostra de domicílio, censo agropecuário e censo demográfico. Todavia, notou-se que os dados utilizados em suas pesquisas não são descritos por meio de metadados padronizados.

Assim também, PS2 relatou que ainda não teve experiência na elaboração de um Plano de preservação de dados de pesquisa. A estratégia de preservação que utiliza é local, e não institucional. Em um procedimento de preservação instintiva, PS2 recomenda a seus orientandos a realizar o armazenamento de dados em nuvem, como pode-se observar:

Não existe um protocolo que eu sigo, é uma coisa meio instintiva de proteger os nossos dados para que não comprometa. Porque imagina, depois de uma pesquisa de uma tese, de um doutorado, três, quatro anos, tu juntando dado e aí acontece uma coisa como essa: dá uma pane no computador e perde tudo! Aí é uma tragédia, né? Não tem como recuperar.

Para PS2, o compartilhamento de dados ocorre no contexto do grupo de pesquisa o qual participa, e, geralmente, em projetos específicos. Além disso, essa prática é influenciada por aspectos pessoais, nos quais o pesquisador avalia a reputação do indivíduo para compartilhar, ou não, seus dados. Nesse sentido, para que o compartilhamento de dados seja exitoso, é necessário haver uma relação de confiança entre os pesquisadores envolvidos. Ainda segundo o entrevistado, o reuso de dados de pesquisa não é um procedimento habitual para ele, uma vez que costuma apenas citar e referenciar trabalhos publicados, e não dados de pesquisa.

Finalmente, ao se referir à gestão de dados abertos em universidades, PS2 descreveu um cenário inseguro, no qual não há garantias de que dados de pesquisa serão protegidos. Isso, em razão de sua preocupação de que instituições façam uso indevido de dados, manipulando-os conforme seus interesses. Referindo-se, especificamente, a dados sensíveis, como dados pessoais, PS2 disse:

Nós estamos vivendo uma situação insólita, porque no mundo hoje tudo isso é big data, dados que são utilizados por grupos, por corporações, para manipular. Nós estamos vivendo um momento, para mim,

dramático, toda discussão que envolve os dados pessoais [...] daqui a pouco, ele [alguém] pega o meu prontuário médico e faz o que quiser, vende aquelas informações para quem quer que seja. Os nossos dados bancários, a nossa vida está exposta, está escancarada. Que tipo de dados são esses? Que tipo de controle nós temos sobre isso? Já como está, já está perigoso!

O Quadro 26 apresenta as dimensões, categorias e propriedades identificadas na entrevista de PS2, por meio de codificação aberta:

Quadro 26 – Codificação aberta da entrevista com PS2: práticas de gestão de dados de pesquisa

Dimensão	Categoria	Propriedade
Discussão sobre Ciência aberta	Diversidade de áreas	Variação de áreas do conhecimento
	Financiamento à pesquisa	Editais públicos
Planejamento	Falta de estímulo ao pesquisador	Dificuldades na obtenção de recursos
Coleta de dados	Técnicas diversas	Métodos quantitativos e qualitativos
Descrição de dados	Padrão local	Descrição estabelecida pelo pesquisador
Preservação de dados	Estratégias locais	Realizada pelos próprios responsáveis da pesquisa
Compartilhamento de dados	Grupo de pesquisa	Projetos de pesquisa específicos
	Reputação de pesquisadores	Relação de confiança
Reuso de dados	Citação de trabalhos publicados	Referência
Gestão de dados abertos em universidades	Cenário inseguro	Insegurança quanto à proteção de dados

Fonte: elaborado pela autora

4.2.1.3 Codificação aberta da entrevista com PS3

De acordo com PS3, a Ciência aberta é um fenômeno ainda pouco discutido nas Ciências Sociais, em parte, devido à indisponibilidade de informações e dados abertos gerados por instituições públicas. Dados esses que tratam, principalmente, de questões sensíveis à sociedade, incluindo registros de crimes, denúncias e processos penais.

PS3 também relatou sua experiência na elaboração de Planos de gestão de dados de pesquisa, ao participar de editais de chamamento público. Nesse sentido, destacou a necessidade da criação de um plano coeso, de modo que qualquer indivíduo que tenha acesso aos dados produzidos possa ser capaz de compreender como os mesmos foram gerados, além de analisar suas potencialidades e limites.

Ainda com relação ao planejamento da pesquisa, PS3 ressaltou que apesar de receber financiamento oriundo de editais públicos, é muito difícil obter recursos, o que restringe o desenvolvimento de estudos.

Para PS3, o treinamento de pesquisadores é o aspecto mais relevante a se tomar nota nos procedimentos de coleta de dados. Na aplicação de questionários, por exemplo, é imprescindível dominar o conteúdo, interagir com os participantes da pesquisa e minimizar a ocorrência de vieses, aumentando assim, a qualidade dos dados. Além disso, os conjuntos de dados coletados são descritos conforme padrões de metadados locais, estabelecidos por PS3.

Ao falar sobre a importância de documentar os procedimentos relacionados às pesquisas que realiza, PS3 ressaltou que isso é essencial para compreender a potencialidade e limite de dados, ou seja, entender os processos nos quais foram submetidos e sob quais contextos foram coletados. Referindo-se à documentação de processos no contexto de uma base de dados, PS3 explicou que dados podem estar aquém de resultados estimados, ou não imprimem qualidade, sendo necessário oferecer detalhes sobre como foram obtidos:

Não adianta você distribuir essa base de dados sem contar para a pessoa que está recebendo, como é que foi todo esse processo. Você pode ter, inclusive, uma imagem enviesada de que esses dados representam muito mais do que de fato eles contêm em si mesmos, ou quais foram as dificuldades que, por sua vez, podem comprometer a representatividade daquela informação, podem comprometer a qualidade daquela informação. Então, eu acho de suma importância descrever, na medida do possível, todos os detalhes relacionados à coleta de dados.

PS3 também informou que não teve experiência na elaboração de um plano de preservação de dados de pesquisa. As estratégias adotadas em relação à preservação de dados são locais, realizadas por PS3, a depender do tipo de dado e de seu formato (digital ou impresso). Por essa razão, há conjuntos de dados armazenados em nuvem, em suportes eletrônicos como, HDs (*Hard Disc*) e arquivos físicos, como foi apontado em sua fala:

Se é possível a disponibilização pública, a gente tenta colocar no site do [grupo de pesquisa], essas informações, e se não é possível a disponibilização pública, ela fica na nuvem e no HD. Então, esse é o máximo de preservação que a gente tem. Uma coisa que a gente faz e que é obrigado por lei também, é guardar todos os questionários das pesquisas pelo prazo de 10 anos [...] se é questionário em papel, a gente tem uma sala, um arquivo morto, que é onde a gente coloca todos esses questionários.

Com relação ao reuso de dados de pesquisa, PS3 afirmou que essa prática ainda é pouco disseminada na área em que atua, por duas razões: 1) a

indisponibilidade de acesso a dados por parte de quem os produziu – inibindo o pensamento de novas questões a partir de velhos dados que não se encontram amplamente disponíveis – 2) a valorização de dados por parte de pesquisadores – refletindo o enaltecimento de seu próprio trabalho de campo, em detrimento de outros trabalhos.

Nas ciências sociais brasileiras, tem uma enorme valorização do seu próprio trabalho de campo [...] assim, tenho alunas que reutilizaram dados que a gente coletou em outros momentos, mas isso não é tão trivial ainda, especialmente, dentro da sociologia.

Além disso, PS3 relatou que ainda faltam orientações sobre a gestão de dados abertos de pesquisa em universidades, sendo necessária a realização de treinamentos para uma real compreensão a respeito do tema. Outro aspecto apontado por PS3 foi a diversidade de práticas em relação a dados de pesquisa, ocasionada por diferenças disciplinares. Por isso, a realização de debates na área das Ciências humanas em relação à disponibilidade de dados é urgente e toca em questões futuras relativas à preservação de memórias.

A partir do Quadro 27 é possível observar as dimensões, categorias e propriedades identificadas na codificação aberta, conforme o relato de PS3.

Quadro 27 – Codificação aberta da entrevista com PS3: práticas de gestão de dados de pesquisa

Dimensão	Categoria	Propriedade
Discussão sobre Ciência aberta	Pouca discussão	Indisponibilidade de dados abertos de pesquisa
	Financiamento à pesquisa	Editais públicos
Planejamento	Falta de estímulo ao pesquisador	Dificuldades na obtenção de recursos
	Plano de gestão de dados	Elaboração de documento
Coleta de dados	Treinamento	Qualidade de dados
Descrição de dados	Padrão local	Descrição estabelecida pelo pesquisador
Documentação de dados	Potencialidade e limite de dados	Visão de processos e contextos
Preservação de dados	Estratégias locais	Realizada pelos responsáveis da pesquisa
Reuso de dados	Indisponibilidade de acesso	Restrições de uso
	Valorização de dados	
Gestão de dados abertos em universidades	Falta de diretrizes	Orientações para a gestão de dados de pesquisa
	Diferenças disciplinares	Diversidade de práticas

Fonte: elaborado pela autora

4.2.1.4 Codificação aberta da entrevista com PS4

De acordo com PS4, há pouca discussão no Brasil sobre o fenômeno da Ciência aberta no campo das Ciências sociais, o que é bastante diferente do cenário internacional, principalmente, em países anglófonos cujos debates já se encontram em desenvolvimento. Para PS4, reflexões sobre esse tópico são essenciais, na medida em que a disponibilização de dados abertos pode implicar em alguns problemas do ponto de vista metodológico, e, ao mesmo tempo, oferecer muitas oportunidades relacionadas a pesquisas colaborativas. Sobre isso, PS4 explicou:

Nas ciências sociais, isso varia de país para país. Mas também existem plataformas, hoje, cada vez mais impulsionadas fora do Brasil, que estão tentando tornar os dados coletados, inclusive dados qualitativos, entrevistas, mesmo de áreas de campo, disponíveis para outros pesquisadores, para que esses dados possam ser analisados por outros pesquisadores também. O que também implica em muitos problemas do ponto de vista metodológico e, ao mesmo tempo, muitas possibilidades de pesquisa conjunta. Então, eu acho que é um debate novo importante que está acontecendo em vários países. Aqui no Brasil, ainda eu vejo muito pouco. Acho que no mundo anglófono, no mundo científico anglófono, isto está mais presente.

Por ser um tema pouco discutido no Brasil, PS4 preocupou-se com o cumprimento de princípios éticos, principalmente, relacionados a acordos de confidencialidade sobre dados de pesquisa. Nessa direção, relatou que costumeiramente lança mão de etnografias para coletar dados e escreve muitos diários de campo, os quais não poderiam ser disponibilizados publicamente sem nenhuma espécie de tratamento técnico. Para tanto, seria necessário, no mínimo, realizar o procedimento de anonimização de dados, de modo a preservar a identificação anônima de indivíduos e não expor informações confidenciais.

Ao falar sobre planejamento de pesquisa, PS4 destacou que participa de uma rede de pesquisa, caracterizada sobretudo pela realização de trabalhos colaborativos que buscam financiamento por meio de editais específicos. Esses, por sua vez, são promovidos por agências de fomento à pesquisa que possuem diferentes orientações em relação à gestão de dados. PS4 também destacou que o trabalho em rede é muito produtivo, uma vez que equipes conectadas em diferentes universidades e centros de pesquisa têm mais acesso a oportunidades de financiamento de projetos. Ademais, disse que já teve experiência na elaboração de planos de gestão de dados, inclusive, como avaliador no contexto de agências de fomento.

Quando coleta dados, PS4 lança mão de diversas estratégias metodológicas. Uma delas é a etnografia, decorrente de observação participante e posterior relato em diários de campo sobre os fenômenos observados. Para ele, tais diários são materiais fundamentais de pesquisa e expressam a visão íntima do pesquisador que o escreveu. Além disso, muitas entrevistas são realizadas, bem como consultas a documentos secundários utilizados para contextualizar os temas em estudo.

Quanto à descrição de conjuntos de dados coletados em diários de campo, PS4 explicou que adota um procedimento local, no qual dados são descritos e organizados em torno de cinco categorias: tempo (cronologias vivenciadas); espaço (cartografias visitadas); descrição densa de situações (como se dá as interações durante a atividade de pesquisa); *insights* analíticos; diário íntimo (autorreflexão do pesquisador em campo de estudo).

No que se refere à documentação de processos relacionados a pesquisas, PS4 comentou que esse procedimento é fundamental para que dados gerados possam ser avaliados quanto ao seu nível de qualidade. Isso inclui a demonstração sobre como dados foram coletados, quais metodologias foram utilizadas e quais conclusões podem ser obtidas. Já em relação ao compartilhamento de dados, PS4 afirmou que disponibiliza com frequência dados de suas pesquisas, mas apenas para seus colegas e participantes de grupo de pesquisa o qual faz parte. Nesse sentido, ainda comentou que os diários de pesquisa que produz necessitam ser anonimizados antes de serem publicados, além de terem uma ordenação sequencial, sem a qual seria muito difícil compreender a narração de fenômenos observados.

Ao abordar aspectos da preservação de dados, PS4 disse que adota estratégias locais, como o armazenamento de pastas de arquivos em computadores ou em pastas protegidas em nuvem. Entretanto, ressaltou que os dados de pesquisa não estariam organizados a ponto de serem disponibilizados publicamente, como seria no contexto de um repositório. Já com relação ao reuso de dados de pesquisa, PS4 comentou que essa prática ocorre em uma escala muito pequena, internamente, nos grupos de pesquisa os quais participa. Assim sendo, já reutilizou dados produzidos por estudantes, bem como esses já reutilizaram dados que gerou.

Ao fazer sugestões para sua universidade, caso a gestão de dados abertos seja institucionalizada, PS4 apontou a necessidade de tratamento e organização de dados por parte da instituição. Tal medida abrange a disponibilização de recursos direcionados à contratação de profissionais capacitados para trabalharem com dados, uma vez que não é possível disseminá-los publicamente sem nenhuma intervenção técnica. Sobre esse aspecto, PS4 explicou:

Que fossem disponibilizados recursos específicos para a organização de dados de pesquisa, editais específicos para a contratação de pesquisadores que pudessem trabalhar para disponibilizar os nossos dados. Porque os dados não é uma questão de você pegar e colocar online, esses dados precisam ser trabalhados para serem disponibilizados. Na minha área, o pesquisador vai ter que ler esses diários de campo e verificar quais são os nomes, quais são os lugares que tem que ser trocados para que isso possa ser disponibilizado. É um material riquíssimo, mas vai demandar muito trabalho para ser disponibilizado [...] por razões éticas, eu tenho o compromisso com os meus interlocutores de não dizer quem eles são.

O Quadro 28 mostra as dimensões, categorias e propriedades identificadas na codificação aberta, a partir do relato de PS4.

Quadro 28 – Codificação aberta da entrevista com PS4: práticas de gestão de dados de pesquisa

Dimensão	Categoria	Propriedade
Discussão sobre Ciência aberta	Cenário internacional	Pouco discutido no Brasil
Aspectos éticos	Acordos de confidencialidade de dados	Anonimização de dados
Planejamento	Financiamento à pesquisa	Editais públicos
	Colaboração científica	Rede de pesquisadores
	Plano de gestão de dados	Elaboração de documento
Coleta de dados	Técnicas diversas	Métodos quantitativos e/ou qualitativos
Descrição de dados	Padrão local	Descrição estabelecida pelo pesquisador
Documentação de dados	Avaliação de dados	Verificação de processos da pesquisa
Compartilhamento de dados	Grupo de pesquisa	Ocorre no escopo do grupo de pesquisa
Preservação de dados	Estratégias locais	Realizada pelos próprios responsáveis da pesquisa
Reuso de dados	Grupos em rede de pesquisa	Colaboração científica
Gestão de dados abertos em universidades	Tratamento de dados	Necessidade

Fonte: elaborado pela autora

4.2.1.5 Codificação aberta da entrevista com PS5

De acordo com PS5, o tópico da Ciência aberta tem sido comentado em associações na área de Sociologia e no programa de pós-graduação no qual atua. Inicialmente, uma questão-chave destacada em seu relato foi a preocupação em observância a aspectos éticos, em especial, a acordos de confidencialidade de dados contidos em termos de sigilo estabelecidos com participantes de pesquisas.

Quanto ao planejamento de pesquisa que conduz, PS5 disse que recebe financiamento de agências de fomento, mediante à participação de editais públicos. Também falou que nunca foi solicitado por nenhuma instituição a elaborar um Plano de gestão de dados.

PS5 comentou que lança mão de diversas estratégias metodológicas para a coleta de dados e que adota seus próprios padrões para descrever conjuntos de dados. Outro aspecto destacado por PS5 foi a importância de documentar os procedimentos relativos à pesquisa, nomeando-os por meio do termo: gerenciamento da prova. Em seu ponto de vista, as provas podem ser entrevistas, documentos, questionários ou quaisquer instrumentos que comprovem que um estudo foi realizado e que fenômenos foram investigados por meio de procedimentos científicos.

No que se refere à prática de compartilhamento de dados, PS5 se reportou à experiência de ter publicado conjuntos de dados em um periódico científico, juntamente com um artigo submetido. É importante salientar que a publicação do artigo não esteve condicionada à disponibilização de dados, que no caso, eram provenientes de entrevistas. Entretanto, poderia ampliar a visibilidade da publicação em níveis internacionais, como pode-se notar em seu relato:

Então, a gente fez um conjunto de entrevistas, se eu não me engano, a gente fez catorze ou quinze, e aí, usamos as entrevistas, montamos um artigo e submetemos a essa revista. No fim das contas, o artigo foi aprovado e a revista, ela está meio de ponta nisso, ela não condicionou a publicação do artigo ao envio das entrevistas, mas ela disse que está investindo nesse projeto, que é importante isso daqui para frente, que ela está no nível internacional. Aí, a gente simplesmente pegou as entrevistas, formatou corretamente e mandou. Então, agora, o artigo já vai sair e as entrevistas, que são a base do artigo, vão estar disponíveis no banco de dados, que também é público.

Ao falar sobre a preservação de dados, PS5 relatou que orienta seus alunos a armazenarem os dados que produzem, mas que nunca elaborou formalmente um Plano de preservação de dados, nem para a sua equipe de pesquisa, nem para o

programa de pós-graduação o qual integra. Quanto à prática de reuso de dados, PS5 disse que não tem o costume de utilizar dados de outras pesquisas, mas sim referenciar conhecimento produzido e publicado. Quando assim o faz, trata-se de trabalho colaborativo que terá como resultado publicações em coautoria. Nesse sentido, PS5 ilustrou sua experiência com o reuso de dados da seguinte maneira:

O que eu cheguei mais próximo foi uma pesquisa que eu fiz, depois um colega fez parecido, aí a gente compartilhou. Nós compartilhamos os dados, então, ele me deu os questionários que ele tinha, eu bolei [criei], ampliei um questionário. Ele aplicou, ampliou o número de questionário, eu peguei algumas entrevistas que eu tinha, ele pegou várias entrevistas que ele tinha e juntos nós constituímos um conjunto de entrevistas e questionários. E aí, nós escrevemos três artigos a partir desses dados: eu autor, ele coautor; eu coautor ele autor; sempre bem claro, bem compartilhado o uso dos dados, a autoria dos artigos, e assim por diante.

O Quadro 29 indica as dimensões, categorias e propriedades assinaladas na codificação aberta, conforme o relato de PS5.

Quadro 29 – Codificação aberta da entrevista com PS5: práticas de gestão de dados de pesquisa

Dimensão	Categoria	Propriedade
Discussão sobre Ciência aberta	Convívio científico	Informações sobre o tema
Aspectos éticos	Acordos de confidencialidade de dados	Anonimização de dados
Planejamento	Financiamento à pesquisa	Editais públicos
Coleta de dados	Técnicas diversas	Métodos quantitativos e/ou qualitativos
Descrição de dados	Padrão local	Descrição estabelecida pelo pesquisador
Documentação de dados	Gerenciamento da prova	Prova de dados produzidos
Compartilhamento de dados	Periódicos científicos	Dados publicados juntamente com artigos
Preservação de dados	Estratégias locais	Realizada pelos próprios responsáveis da pesquisa
Reuso de dados	Citação de trabalhos publicados	Referência
	Colaboração científica	Trabalhos em coautoria

Fonte: elaborado pela autora

4.2.1.6 Codificação aberta da entrevista com PS6

De acordo com PS6, o tópico da Ciência aberta tem sido pouco discutido em nível institucional, não representando ainda, uma questão departamental. Com relação ao planejamento das pesquisas que realiza, comentou que participa de editais de financiamento demandados, principalmente, por instituições governamentais, como

ministérios e secretarias. Pode-se falar de pesquisas indutoras de políticas públicas. Entretanto, PS6 destacou a dificuldade na obtenção de recursos, uma vez que mencionou uma crise gigantesca para a produção científica no contexto brasileiro.

PS6 também relatou um pouco de sua experiência na elaboração de um Plano de gestão de dados de pesquisa direcionado à polícia civil de determinado estado para obter bancos de dados relacionados aos índices de criminalidade. Para tanto, foi preciso lançar mão de argumentos da Lei de acesso à informação e descrever todo o planejamento relativo aos dados, incluindo os critérios de coleta e sua destinação.

Quanto à coleta de dados, PS6 demonstrou a diversidade de procedimentos metodológicos adotados nas pesquisas que realiza. Nessa direção, ilustrou que busca caracterizar o estado da arte sobre determinados temas, por meio de revisões integrativas e de estudos bibliométricos. Semelhantemente, realiza entrevistas, grupos focais e escreve diários de campo, como pode-se observar em seu relato:

O grande volume são dados qualitativos, então, a gente tem muita coisa, transcrições de entrevistas, de grupos focais, que são as duas técnicas que a gente tem mais usado, né? Para fazer pesquisa com esses públicos, além de diários de campo, né? Quando você está fazendo as observações empíricas e tal, então, a gente, no caso das entrevistas e dos grupos focais, a gente tem o roteiro, o roteiro guia e a transcrição literal daquele material. Esse é o tipo de informação que a gente produz.

Outro aspecto destacado por PS6 foi a coleta de dados em meio a canais difusos, cujas informações encontram-se descentralizadas. Por conseguinte, foi sugerida a criação de repositórios institucionais que disponibilizem bancos de dados padronizados e interoperáveis.

A descrição de conjuntos de dados é realizada localmente por PS6, que por meio da criação de planilhas eletrônicas, atribui metadados, como: localidade; gênero; raça; local de moradia; profissão; evento (geralmente relacionado à tipificação penal). Além disso, PS6 comentou o quanto a documentação de processos é relevante para o aprimoramento de pesquisas, na medida em que possibilita a complementação de estudos e o avanço de contrapropostas àquilo que se definiu.

Com relação à preservação de dados, PS6 disse que adota estratégias locais, como o armazenamento de dados em nuvem, porém, nada institucionalmente sistematizado. Até então, declarou que nunca elaborou um Plano de preservação de dados. Semelhantemente, disse que ainda não tinha experiência no compartilhamento

de dados abertos. Todavia, expressou o seu anseio em fazê-lo por meio da criação de um portal/repositório de acesso público.

Em se tratando de reuso de dados de pesquisa, PS6 afirmou que faz uso de dados disponibilizados publicamente por algumas instituições, tais como o Instituto de Pesquisa Econômica Aplicada (Ipea), Fundo das Nações Unidas para a Infância (Unicef) e Organização das Nações Unidas para a Educação, a Ciência e a Cultura (Unesco). Os dados abordam, principalmente, aspectos socioeconômicos relacionados a moradia, infância, adolescência e escolaridade.

PS6 faz algumas reflexões no contexto da universidade em que atua, caso a gestão de dados abertos seja institucionalizada, as quais pode-se destacar três: 1) a necessidade de discussões sobre o tema; 2) a importância das diferenças disciplinares; 3) receio de pesquisadores em compartilhar dados.

Sobre a necessidade de discussões sobre o tema, PS6 sugeriu intensificar esse diálogo entre departamentos e programas de pós-graduação da universidade:

É que a gente precisa falar disso de modo interdepartamental, né? Ou começar a partir de. Às vezes, envolver os departamentos é muito difícil, mas envolver os programas de pós-graduação pode ser um caminho. Então, eu ainda não vi nesse momento agora [...] a gente precisa conhecer para compreender. Eu preciso conhecer sobre esse tema, eu professora, eu aluna, para poder me ver planejando a gestão de dados.

Considerando as diferenças disciplinares no contexto universitário, PS6 explicou que as áreas de conhecimento possuem especificidades na produção e disseminação científica, as quais devem ser discutidas no âmbito dos dados de pesquisa:

Porque há coisas específicas, eu acho que dentro das humanidades há alguns dramas que às vezes não se colocam nas Ciências duras, entendeu? Que devem ter outras questões em relação aos dados. Então, a minha primeira sugestão para [nome da universidade] é promover o ciclo de seminários, debates, capacitações, workshop, sei lá, o nome que a gente queira dar, nos agregando por grandes áreas de conhecimento [...] porque cada programa de pós-graduação vai irradiar para seu departamento as discussões, né? E eu acho que esses encontros por grandes áreas têm que, necessariamente, envolver discentes, não pode só envolver professores.

Sobre o receio de pesquisadores em compartilhar os dados que produzem, PS6 sugeriu, mais uma vez, a realização de conversas sistemáticas acerca do tema que incluam perspectivas de especialistas, pesquisadores e estudantes:

Só quem estuda, só alguma especialista da informação está realmente atenta a essa questão. Então, para a gente poder romper uma questão de medo, né? Porque quando eu li, eu fiquei com medo: “ai, eu vou botar nossos dados? Ai meu Deus! O que eles vão fazer?” A gente fica com os receios que são puro mito. Mas para que a gente não incorra na ideia do medo, do temor sobre a publicitação que vai ampliar muito a vida dos pesquisadores, a gente precisa fazer conversas sistemáticas sobre esse assunto, entendeu? Por especialistas ouvindo as demandas das áreas e tentando compor um conjunto de

recomendações necessárias. É um pouco o meu raciocínio, mas não pode abrir mão de discentes, senão a gente não sai do lugar.

No Quadro 30 pode-se ver as dimensões, categorias e propriedades identificadas na codificação aberta, a partir do relato de PS6.

Quadro 30 – Codificação aberta da entrevista com PS6: práticas de gestão de dados de pesquisa

Dimensão	Categoria	Propriedade
Discussão sobre Ciência aberta	Pouca discussão	Nível institucional
	Financiamento à pesquisa	Editais públicos
Planejamento	Falta de estímulo ao pesquisador	Dificuldades na obtenção de recursos
	Plano de gestão de dados	Elaboração de documento
Coleta de dados	Técnicas diversas	Métodos quantitativos e/ou qualitativos
	Dados difusos	Dados dispersos em instituições
Descrição de dados	Padrão local	Descrição estabelecida pelo pesquisador
Documentação de dados	Aprimoramento de pesquisas	Descrição de processos na pesquisa
Preservação de dados	Estratégias locais	Realizada pelos próprios responsáveis da pesquisa
Compartilhamento de dados	Portal/repositório	Necessidade de canais de disponibilização de dados
Reuso de dados	Dados institucionais	Fontes institucionais
	Debates	Discussão sobre o tema
Gestão de dados abertos em universidades	Diferenças disciplinares	Práticas relacionadas aos dados de pesquisa
	Receio de pesquisadores	Pouco conhecimento sobre o tema

Fonte: elaborado pela autora

4.2.1.7 Codificação aberta da entrevista com PS7

Segundo PS7, as discussões sobre Ciência aberta ocorrem em nível informal, no departamento da universidade na qual faz parte. Apesar de achar um tema relevante, reconheceu que ainda é pouco trabalhado.

Sobre o planejamento de pesquisas que conduz, PS7 disse que elabora projetos que são enviados a agências de fomento para concorrer a financiamentos. Relatou, ainda, que se preocupa com a questão dos dados que produz, porém, considerou

que um Plano de gestão de dados era uma expressão muito significativa para abranger o tipo de organização que realiza.

Em seus trabalhos, PS7 comentou que lança mão de diversas estratégias de coleta de dados, sobretudo, a realização de entrevistas em um contexto denominado: bola de neve, no qual desenvolve-se uma rede de possíveis entrevistados. Porém, os conjuntos de dados obtidos são descritos localmente, sem a utilização de metadados padronizados.

Quanto à documentação de processos relacionados à pesquisa, PS7 argumentou que tal procedimento possibilita a comprovação da veracidade de dados, quando esses são expostos à avaliação. Além disso, relatou que documenta processos relacionados aos dados que produz, mesmo não sendo auditado por alguma instituição.

Sobre compartilhamento de dados de pesquisa, PS7 disse que procura disponibilizá-los abertamente, em especial, quando se trata de dados quantitativos. No que concerne a dados qualitativos, comentou que tal abertura poderia gerar interpretações equivocadas de dados, e, conseqüentemente, má utilização. Já quanto à preservação de dados, PS7 relatou que ainda não teve experiência na elaboração de um Plano de preservação de dados, mas que adota estratégias próprias de preservação, como o armazenamento de entrevistas, transcrições, vídeos, e áudios em um *Drive* de computador.

Conforme explicou PS7, o reuso de dados ocorre no âmbito do grupo de pesquisa o qual integra, sendo muito comum o intercâmbio de dados com orientandos. Por outro lado, não costuma reutilizar dados de estudos os quais não foram supervisionados por ele.

Finalmente, ao falar sobre a gestão de dados de pesquisa no contexto de universidades, PS7 sugeriu que esse tópico seja amplamente discutido, para ser coletivamente construído pela comunidade acadêmica, como expôs em seu comentário:

A sugestão central é a de fazer uma coisa amplamente discutida, antes de ser implementada. Que seja uma construção coletiva do próprio coletivo de professores e professoras da universidade, para que não seja uma coisa de cima para baixo. Então, que a comunidade acadêmica participe para que haja uma elaboração que seja mais interessante para o conjunto e para a própria comunidade acadêmica.

O Quadro 31 mostra as dimensões, categorias e propriedades identificadas na codificação aberta, com base na entrevista de PS7.

Quadro 31 – Codificação aberta da entrevista com PS7: práticas de gestão de dados de pesquisa

Dimensão	Categoria	Propriedade
Discussão sobre Ciência aberta	Entre pesquisadores	Discussão informal, em menor escala
Planejamento	Financiamento à pesquisa	Editais públicos
Coleta de dados	Técnicas diversas	Métodos qualitativos
Descrição de dados	Padrão local	Descrição estabelecida pelo pesquisador
Documentação	Avaliação de dados	Verificação de processos da pesquisa
Preservação de dados	Estratégias locais	Realizada pelos próprios responsáveis da pesquisa
Compartilhamento de dados	Interpretação equivocada de dados	Uso inadequado
Reuso de dados	Grupos em rede de pesquisa	Colaboração científica
Gestão de dados abertos em universidades	Debates	Discussão sobre o tema

Fonte: elaborado pela autora

4.2.1.8 Codificação aberta da entrevista com PS8

De acordo com PS8, o fenômeno da ciência aberta atingiu o alcance internacional, na medida em que alguns editores têm solicitado a autores a disponibilização de dados como condição para a publicação de artigos. Na perspectiva de PS8, essa é uma tendência no meio científico que possibilita a replicação de estudos, e a descrição detalhada de rotinas metodológicas e de análises realizadas.

Com relação ao planejamento de suas pesquisas, PS8 relatou que costuma utilizar dados secundários, em razão de dificuldades na obtenção de recursos para a produção de novos dados. Ao falar sobre a estrutura de trabalho que dispõe na universidade a qual faz parte, PS8 expõe a necessidade de melhorias na disponibilização de recursos aplicados à pesquisa:

Eu levava o meu laptop, e a Internet acessava pelo meu celular. Eu colocava ele como um ponto de roteador e usava porque a estrutura é muito ruim. Para você ter uma ideia, a cadeira que eu sento no meu gabinete, eu comprei e levei, porque a que estava lá estava caindo aos pedaços, a mesa é horrível, toda. Então, você não tem estrutura, pelo menos no departamento de sociologia da [universidade]. A estrutura é muito ruim, aí você vai com o seu dinheiro montando alguma coisa, se

you want to stay there. From there, you think about having a place where you deposit your research, where people will have access, etc., this is a dream!

PS8 relatou que ainda não teve experiência na elaboração de um Plano de gestão de dados. Quanto à coleta de dados, há o predomínio daqueles oriundos de fontes secundárias, de bases alimentadas por instituições como o Instituto Brasileiro de Geografia e Estatística (IBGE) e o departamento de informática do Sistema Único de Saúde do Brasil (DATASUS). Isso, principalmente, em razão da diminuição de custos para a obtenção de dados de pesquisa.

Os conjuntos de dados são descritos por PS8, não havendo a adoção de padrões internacionais de descrição. Além disso, PS8 considerou que a documentação relacionada aos dados é fundamental para verificação de processos da pesquisa, facilitando a avaliação de dados e a identificação de inconsistências metodológicas. No que se refere ao compartilhamento de dados, percebe-se que essa prática ocorre no escopo do grupo de pesquisa o qual integra PS8.

No que diz respeito à preservação de dados, PS8 comentou que ainda não elaborou um Plano de preservação de dados de pesquisa. Tal preservação ocorre sob sua responsabilidade, por meio de armazenamento local, como se pode observar:

This preservation of data was done by the people themselves, you left it in an HD, in a drawer and ready! You didn't have any, so there I can say, these levels of security where this will be stored, in short, no!

PS8 disse que não percebia preconceito em relação ao reuso de dados de pesquisa na Sociologia. Entretanto, comentou sobre a importância de se avaliar a qualidade de dados a serem reutilizados, principalmente no que se refere à autoria de quem os produziu, ao nível de estruturação em que se encontram e às condições de acesso os quais estão disponíveis. A partir de sua experiência de reuso em bases de dados do Instituto Brasileiro de Geografia e Estatística (IBGE), PS8 explicou sobre a preocupação em utilizar dados consolidados:

People use consolidated data, for example, the census from IBGE. IBGE makes this available, and it is available on a platform that you find this data bank. So, you don't have this concern about quality, you can analyze some things, in short, but they are consolidated data. You have international data that people use, the whole world uses, so, you have a greater tranquility to use this information.

Por fim, ao falar da gestão de dados abertos em universidades, PS8 explicou que essa temática necessita ser debatida nas instituições, de modo que se discutam

aspectos como: custos envolvidos; plataformas a serem adotadas; políticas de acesso e uso de dados; critérios para a disponibilização de banco de dados.

O Quadro 32 mostra as dimensões, categoria e propriedades identificadas na codificação aberta, com base na entrevista de PS8.

Quadro 32 – Codificação aberta da entrevista com PS8: práticas de gestão de dados de pesquisa

Dimensão	Categoria	Propriedade
Discussão sobre Ciência aberta	Cenário internacional	Tendência
Planejamento	Falta de estímulo ao pesquisador	Dificuldades na obtenção de recursos
Coleta de dados	Dados secundários	Fontes secundárias
Descrição de dados	Padrão local	Descrição estabelecida pelo pesquisador
Documentação de dados	Avaliação de dados	Verificação de processos da pesquisa
Compartilhamento de dados	Grupo de pesquisa	Ocorre no escopo do grupo de pesquisa
Preservação de dados	Estratégias locais	Realizada pelos próprios responsáveis da pesquisa
Reuso de dados	Qualidade de dados	Verificação da qualidade
Gestão de dados abertos em universidades	Debates	Necessidade de discussão do tema

Fonte: elaborado pela autora

4.2.1.9 Codificação aberta da entrevista com PS9

De acordo com PS9, a Ciência aberta é um fenômeno ainda pouco discutido na área em que atua. O seu contato com a temática ocorreu com pesquisadores do departamento de informática de sua universidade e na participação de eventos sobre Ciência de dados. Nessa direção, ressaltou que esse não é um tópico de discussão consolidado nas Ciências Sociais.

Ao abordar aspectos do planejamento das pesquisas que conduz, PS9 comentou sobre a falta de estímulo ao pesquisador, uma vez que observa uma crise de contingenciamento de recursos destinados à produção científica. Além disso, relatou que ainda não teve experiência na elaboração de um plano de gestão de dados de pesquisa.

Já quanto à coleta de dados, PS9 relatou que lança mão de diversas técnicas com essa finalidade. Assim, disse que utiliza dados governamentais, predominantemente, bem como realiza *surveys* e entrevistas. Ademais, os dados obtidos recebem uma descrição local, com a inclusão de metadados básicos como título, autor, resumo e palavras-chave.

Para PS9, o compartilhamento de dados não é um procedimento corriqueiro. Relatou ter tido algumas experiências em projetos específicos, cujo compartilhamento ocorreu, predominantemente, no grupo de pesquisa o qual integra. No que se refere à documentação relativa aos dados que produz, PS9 ressaltou que apesar de relevante, trata-se de uma preocupação particular, não institucionalizada. Sobre isso, PS9 explicou:

Isso teria que ser consensual, teria que ter uma prática coletiva, uma prática, digamos, institucionalizada nisso, né? Então, acabam sendo práticas individuais, não é? [...] então, eu acho que é importante, mas tem que ser levado de forma coletiva, de forma institucional, mesmo porque a preocupação maior no final do projeto é você prestar conta do projeto.

Quanto à preservação de dados, PS9 afirmou que ainda não teve experiência na elaboração de um Plano de preservação de dados de pesquisa. Essa preservação ocorre rotineiramente, por meio de armazenamento de dados em discos rígidos portáteis, *pen drives* e em nuvem. São diferentes mídias que acabam suscetíveis à degradação ambiental e temporal, em razão da falta de migração para suportes mais atualizados. A respeito disso, PS9 relatou:

Então, tem esse material de entrevistas, por exemplo, em áudios e também transcritas e digitalizadas, né? Que estão gravadas nesse tipo de mídia, né? Ou seja, na verdade, são discos e, às vezes, a gente vai perdendo, né? Quer dizer, por exemplo, tem disco lá que daqui a pouco vou ter que usar o aplicativo porque a interface já não bate, são esses problemas que os museus enfrentam, que os arquivos enfrentam, que os grandes arquivos enfrentam e a gente não tem muita, não tem capacidade técnica de lidar com isso.

Quanto ao reuso de dados, PS9 disse que recomenda a seus orientandos que acessem plataformas de consórcios de informação na área de Ciências sociais para reutilizarem bancos de dados disponibilizados por outros pesquisadores. Além disso, comentou que reutiliza muitos dados institucionais *online*, acessados no sítio da Fundação Getúlio Vargas (FGV).

Considerando o exposto, PS9 sugeriu a intensificação de debates no contexto universitário para que a gestão de dados abertos seja implementada. Nessa direção,

PS9 destacou a importância da divulgação sobre a temática por meio de eventos e da ministração de disciplinas em módulo livre:

Para a faculdade? Eu acho que a gente teria que ter, ou assim, fazer eventos. Vamos dizer assim: encorajar, né? Convidar algum professor para dar uma disciplina ou convidar os alunos para assistirem disciplinas de núcleo livre em outras unidades.

O Quadro 33 mostra as dimensões, categoria e propriedades identificadas na codificação aberta, a partir da entrevista com PS9.

Quadro 33 – Codificação aberta da entrevista com PS9: práticas de gestão de dados de pesquisa

Dimensão	Categoria	Propriedade
Discussão sobre Ciência aberta	Pouca discussão	Discussão entre departamentos
Planejamento	Falta de estímulo ao pesquisador	Dificuldades na obtenção de recursos
Coleta de dados	Técnicas diversas	Métodos quali e quantitativos
Descrição de dados	Padrão local	Descrição estabelecida pelo pesquisador
Documentação de dados	Estratégias individuais	Preocupação particular, não-institucional
Compartilhamento de dados	Projeto de pesquisa	Grupo de pesquisa
Preservação de dados	Estratégias locais	Realizada pelos próprios responsáveis da pesquisa
Reuso de dados	Consórcios de informação	Acesso a banco de dados
Gestão de dados abertos em universidades	Debates	Divulgação na universidade

Fonte: elaborado pela autora

4.2.1.10 Codificação aberta da entrevista com PS10

Para PS10, a Ciência aberta é uma novidade, um tópico no qual ainda não se vê um debate explicitado, apesar de haver o compartilhamento de dados e informações na área. Nesse contexto, o aspecto ético foi logo destacado por PS10, uma vez que foi demonstrado o compromisso de não identificar participantes de pesquisas, preservando-os de quaisquer exposições. Sobre isso, PS10 comentou:

Nós trabalhamos com pessoas, o nosso estudo não é em um laboratório, não é trabalhando com reagentes e tal [...] sempre assumimos o compromisso com esses nossos entrevistados, esses nossos informantes, de que nós vamos manter o anonimato e que nós não vamos divulgar nenhuma informação que possa, eventualmente, gerar problema para eles.

No que se refere ao planejamento de pesquisas que conduz, PS10 relatou que participa, recorrentemente, de editais para obter financiamento e que ainda não teve experiência na elaboração de um Plano de gestão de dados de pesquisa. Já quanto à coleta de dados, observou-se a utilização de diversas técnicas na obtenção de dados quali ou quantitativos. Nessa direção, PS10 destacou a aplicação de questionários e a realização de entrevistas.

PS10 disse que não adota um padrão de descrição para os conjuntos de dados que manipula, apenas realiza uma caracterização local de dados, que se dá por meio de tabelas no *Excel*. Além disso, PS10 comentou que considera importante a documentação de processos relacionados a dados, por intermédio de registros no laboratório de pesquisa. Isso, em razão da alta rotatividade da equipe de pesquisa, composta principalmente por mestrandos e doutorandos. Sob tal perspectiva, o processo da documentação é fundamental para a disseminação de informações que tramitam no laboratório de pesquisa, como se pode notar:

Então, em um primeiro aspecto que eu procuro orientar os mestrandos, doutorandos e os alunos de graduação, também, é sempre procurar aprovar o projeto de pesquisa junto ao comitê de ética. Então, esse, eu acho que já seria um primeiro nível de documentação. O outro nível é exatamente esse: sempre que for questionário, que isso seja tabulado e disponível no Google drive do nosso laboratório, para que todos saibam que estão ali os dados e, se tiverem interesse, possam utilizar. Ou então, mesmo que não tenha ninguém neste momento da pesquisa que tenha interesse, em uma pesquisa futura, [alguém] possa retomar os dados pré-analisados. Como eu te falei, é muito raro que isso aconteça no nosso grupo de pesquisa. Como a gente trabalha muito com mestrandos e doutorandos, você tem uma rotatividade muito grande, você tem mudanças de temas de pesquisas, mas esses dados ficam registrados nos nossos laboratórios.

Dessa forma, depreende-se que o compartilhamento de dados ocorre em nível local, dentre integrantes do mesmo grupo de pesquisa. Além disso, PS10 comentou que ainda não possui experiência em disponibilizar os dados de suas pesquisas de forma amplamente aberta.

Ao falar sobre preservação de dados de pesquisa, PS10 disse que ainda não elaborou nenhum Plano de preservação de dados de pesquisa e que lança mão de estratégias as quais denominou “artesanal” para o armazenamento de dados. Em síntese, tal processo se dá por meio da disponibilização de dados no *google drive* do laboratório. Quanto ao reuso de dados, PS10 afirmou que essa não é uma prática comum na área em que atua, visto que costuma citar trabalhos publicados sobre o mesmo tema ou produzidos pelo grupo de pesquisa que integra.

Ao abordar aspectos da gestão de dados abertos em universidades, PS10 narrou algumas das dificuldades vivenciadas por pesquisadores que coordenam projetos de pesquisa, evidenciando a sobrecarga desses profissionais e a falta de infraestrutura na produção científica. Assim sendo, PS10 sugeriu a intervenção de equipes que orientem e viabilizem a gestão de dados em nível institucional, conforme explicou:

Eu acho que a sugestão principal que eu daria é que não se dê mais trabalho para o pesquisador e para os integrantes do seu grupo de pesquisa. Eu acho que a gente faz pesquisa no Brasil de uma forma extremamente amadorística, sem infraestrutura, sem condições. O professor coordenador do projeto, ele acaba ficando responsável da gestão dos recursos financeiros, da gestão científica, da pesquisa, da coordenação do grupo de pesquisa, da produção de resultados, e se ele ainda tiver que ir fazer, não ele diretamente, mas pessoas do seu grupo, fica extremamente sobrecarregado [...] então assim, voltando para a sua questão, a sugestão é essa: que se for fazer, que eu acho positivo, é que tenha equipes, pessoas, funcionários, técnicos que possam orientar e fazer boa parte desse trabalho.

O Quadro 34 apresenta as dimensões, categorias e propriedades identificadas na codificação aberta, a partir da entrevista com PS10.

Quadro 34 – Codificação aberta da entrevista com PS10: práticas de gestão de dados de pesquisa

Dimensão	Categoria	Propriedade
Discussão sobre Ciência aberta	Pouca discussão	Tema recente
Aspectos éticos	Acordos de confidencialidade de dados	Preocupações éticas
Planejamento	Financiamento à pesquisa	Editais públicos
Coleta de dados	Técnicas diversas	Métodos quali e quantitativos
Descrição de dados	Padrão local	Descrição estabelecida pelo pesquisador
Documentação de dados	Rotatividade de pesquisadores	Registros no laboratório
Compartilhamento de dados	Grupo de pesquisa	Ocorre no escopo do grupo de pesquisa
Preservação de dados	Estratégias locais	Realizada pelos responsáveis da pesquisa
Reuso de dados	Citação de trabalhos publicados	Referências
Gestão de dados abertos em universidades	Sobrecarga de pesquisadores	Falta de estrutura

Fonte: elaborado pela autora

4.2.1.11 Síntese da codificação aberta

O processo de codificação aberta nas entrevistas resultou em uma primeira descrição das práticas de gestão de dados utilizadas por pesquisadores no campo da Sociologia. Nesse sentido, foram identificadas 45 categorias, dispostas em dez dimensões. As categorias abertas identificadas passaram por processo de codificação axial, convertendo-se em outras dez categorias axiais, que, por sua vez, se desdobraram em doze subcategorias, as quais podem ser observadas no Quadro 35.

Quadro 35 – Síntese da codificação aberta – Sociologia: práticas de gestão de dados de pesquisa

Dimensão	Categoria aberta	Categoria axial	Subcategoria axial
Discussão sobre Ciência aberta	Entre pesquisadores	Ciência aberta	Convívio acadêmico
	Diversidade de áreas		
	Pouca discussão		
	Cenário internacional		
Aspectos éticos	Convívio científico	Ética na pesquisa	Acordos de confidencialidade de dados
	Acordos de confidencialidade de dados		
	Indicação de responsabilidade de dados		
Planejamento	Publicação antecipada dados	Planejamento da pesquisa	Financiamento à pesquisa
	Custos		
	Falta de estímulo ao pesquisador		
	Plano de gestão de dados		
	Financiamento à pesquisa		
Coleta de dados	Colaboração científica	Coleta de dados	Diversidade de dados
	Técnicas diversas		
	Dados difusos		
Descrição de dados	Dados secundários	Descrição de dados	Padrão local
	Padrão local		
Documentação de dados	Histórico de pesquisa	Documentação de dados	Histórico da pesquisa
	Potencialidade e limite de dados		
	Aprimoramento de pesquisas		
	Avaliação de dados		
	Gerenciamento da prova		
	Estratégias individuais		
Rotatividade de pesquisadores			

Compartilhamento de dados	Grupo de pesquisa	Compartilhamento de dados	Grupo de pesquisa
	Instituições de fomento à pesquisa		
	Insuficiência de recursos		
	Portal/repositório		
	Interpretação equivocada de dados		
Preservação de dados	Projeto de pesquisa	Preservação de dados	Estratégias locais
	Estratégias locais		
Reuso de dados	Grupos em rede de pesquisa	Reuso de dados	Redes de pesquisa
	Indisponibilidade de acesso		
	Valorização de dados		
	Dados institucionais		
	Qualidade de dados		
	Consórcios de informação		
	Citação de trabalhos publicados		
Gestão de dados abertos em universidades	Insuficiência de recursos	Gestão de dados abertos em universidades	Debates sobre o tema
	Falta de diretrizes		
	Tratamento de dados		
	Debates		
	Diferenças disciplinares		Diferenças disciplinares
	Receio de pesquisadores		
Sobrecarga de pesquisadores			

Fonte: elaborado pela autora

4.2.2 Codificação axial - Sociologia

Esta seção relacionará as categorias e subcategorias identificadas a partir dos dados obtidos no estudo, com base no processo de codificação axial realizado. Sobre esse processo, Strauss e Corbin (2008, p. 123-124) explicam:

Codificação axial: o processo de relacionar categorias às suas subcategorias, é chamado de “axial” porque ocorre em torno do eixo de uma categoria, associando categorias ao nível de propriedades e dimensões [...] o objetivo da codificação axial é começar o processo de reagrupamento dos dados que foram divididos durante a codificação aberta. Na codificação axial, as categorias são relacionadas às suas subcategorias para gerar explicações mais precisas e completas sobre os fenômenos (STRAUSS; CORBIN, 2008, p. 123-124).

4.2.2.1 Categoria “Ciência aberta”

Conforme o entendimento de Cribb e Hartomo (2010), a Ciência aberta é um fenômeno que argumenta em favor do compartilhamento do conhecimento científico produzido, de modo que esse seja disponibilizado sem restrições de acesso e em linguagem compreensível. Uma vez que a produção científica demanda esforço intelectual, dinheiro, tempo, dentre outros recursos, o conhecimento disseminado pode ser considerado insumo para a resolução de desafios complexos vivenciados na atualidade, os quais poderiam estar relacionados, por exemplo, à cura de doenças e prevenção de desastres ambientais.

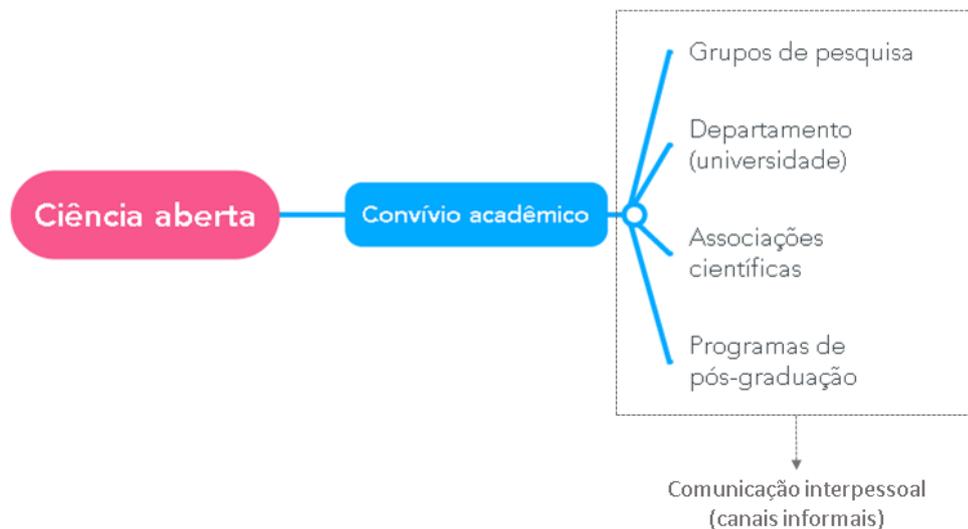
A análise dos dados obtidos nas entrevistas realizadas com pesquisadores da área de Sociologia permitiu observar que, apesar de relevante, o tópico da Ciência aberta ainda é pouco discutido nesse contexto. Conforme apontaram PS3, PS6, PS9 e PS10, cientistas compartilham dados, informações e geram novos conhecimentos rotineiramente. Entretanto, até o momento, não há um debate explícito sobre o tema entre pesquisadores da área e que seja conduzido em nível institucional.

Nesse sentido, os dados resultantes da pesquisa apontam que o acesso à informação sobre Ciência aberta decorre, principalmente, do convívio acadêmico entre os pesquisadores na área de Sociologia, o que inclui discussões em grupos de pesquisa, no departamento da instituição em que atuam, nas associações em que

são membros e/ou nos programas de pós-graduação os quais estão vinculados. Apesar de não ser meio exclusivo de comunicação nesse contexto, destaca-se a importância dos canais informais para o intercâmbio de informações a respeito da Ciência aberta.

Com base no exposto, a categoria “Ciência aberta” é influenciada pela comunicação que ocorre no âmbito do convívio entre pesquisadores. Por essa razão, tal categoria está relacionada à subcategoria “convívio acadêmico” e abrange discussões no âmbito de grupos de pesquisa, departamentos universitários, associações científicas e programas de pós-graduação, tal como é representado na Figura 20.

Figura 20 – Categoria Ciência aberta - Sociologia



Fonte: elaborado pela autora

- Subcategoria “Convívio acadêmico”

Com base nos relatos obtidos na pesquisa, observou-se que o convívio acadêmico entre pesquisadores contribui para a divulgação de informações relativas à Ciência aberta e amplia a discussão sobre o tema. Nesse contexto, o convívio acadêmico pode ser entendido como o conjunto de experiências trocadas por pesquisadores, em decorrência do próprio meio acadêmico e da comunicação científica em termos formais e, especialmente, informais.

Sendo assim, pesquisadores entrevistados afirmaram que a Ciência aberta ainda não é uma pauta abordada em eventos de grande visibilidade no campo das Ciências

Sociais. As discussões também parecem incipientes em nível institucional, apesar de ocorrerem no âmbito de associações na área de Sociologia, programas de pós-graduação em que atuam e eventos científicos. PS1 fez o seguinte comentário a respeito disso:

Existe sim, não uma discussão propriamente pública, isso eu não vejo sendo pautado de maneira repetida em eventos de maior porte, tipo congresso da Sociedade Brasileira de Sociologia, congressos maiores, a ISA [International Sociological Association], por exemplo, International Sociological Society. Eu não tenho visto isso aparecer, mas tenho visto isso aparecer muito entre pesquisadores.

Outro aspecto observado em entrevistas realizadas foi que a Ciência aberta tem sido comentada entre pesquisadores em contexto informal, durante os afazeres rotineiros da pesquisa e ensino. Por meio de comunicação interpessoal, tais pesquisadores trocam informações sobre o referido tópico, principalmente, no departamento da universidade em que atuam, no grupo de pesquisa que integram e com os estudantes em sala de aula.

Com base nos relatos dos entrevistados, ressalta-se o papel fundamental de canais informais na comunicação científica para a reflexão de temas atuais, que ainda não se encontram amplamente contextualizados na área. Logo, pode-se dizer que a comunicação informal é o ponto de partida para a obtenção de informações a respeito desse tema, além de possibilitar a atualização de pesquisadores sobre o mesmo.

Ao compararem o estágio de discussões sobre Ciência aberta no Brasil com o de outros países, principalmente anglófonos, entrevistados (PQ4 e PQ8) disseram que o contexto brasileiro ainda está se desenvolvendo em direção a um consenso na área. Logo, infere-se a existência de um cenário incipiente caracterizado, sobretudo, por discussões em nível informal e não institucionalizado. Tal comparação é visível, na medida em que, internacionalmente, editores científicos têm requisitado a autores, cada vez mais, a disponibilização de dados como requisito para a publicação de artigos.

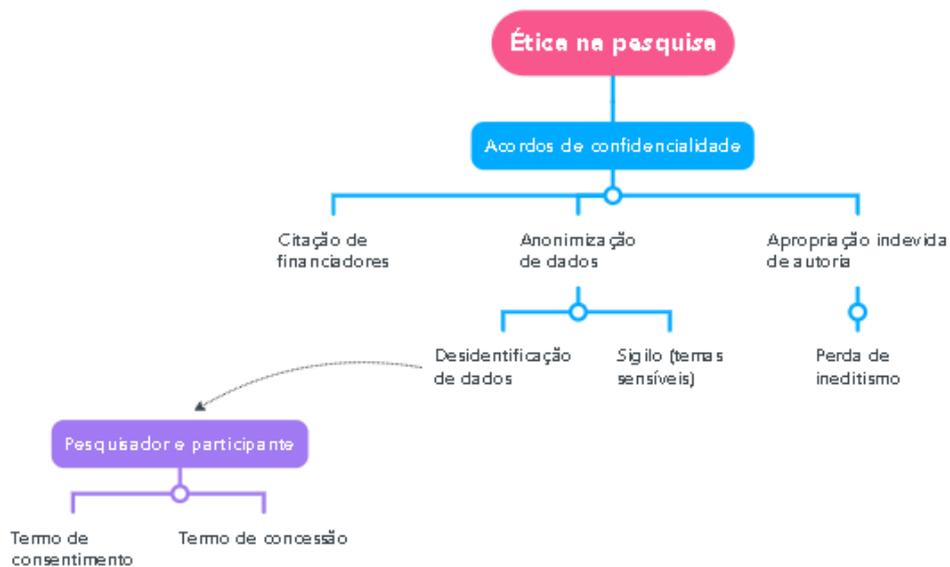
4.2.2.2 Categoria “Ética na pesquisa”

Os dados obtidos no estudo demonstraram que a categoria “Ética na pesquisa”

está diretamente relacionada ao estabelecimento de acordos de confidencialidade sobre dados. Tais acordos são tão importantes nesse contexto, que constituíram uma subcategoria referente à ética na pesquisa.

Como demonstra a Figura 21, os acordos de confidencialidade de dados abrangem a devida citação de instituições financiadoras de pesquisas, a anonimização de dados e procuram evitar a apropriação indevida de dados. Nesses processos, é interessante notar que o ato de anonimizar dados inclui procedimentos de desidentificação de dados e de precauções relacionadas ao sigilo de dados que tratam sobre temas sensíveis. Tais procedimentos referem-se a termos estabelecidos entre pesquisadores e participantes de pesquisas a respeito do consentimento da geração de dados, bem como da concessão de disponibilização de dados. Sob outro ângulo, a apropriação indevida de dados pode ocasionar na perda de ineditismo de dados, e, conseqüentemente, de trabalhos científicos.

Figura 21 – Categoria Ética na pesquisa - Sociologia



Fonte: elaborado pela autora

- Subcategoria “Acordos de confidencialidade de dados”

Conforme relatado por pesquisadores entrevistados (PS1, PS4 e PS5), pode-se notar o quão relevante são a definição e o cumprimento de acordos de

confidencialidade de dados no contexto da pesquisa. Com relação a tais acordos destacam-se três aspectos: 1) devida citação de financiadores da pesquisa (desde que não se trate de dados confidenciais, os quais não poderiam ser disponibilizados abertamente); 2) anonimização de participantes; 3) apropriação indevida de ideias e consequente perda de ineditismo.

Geralmente, quando o pesquisador recebe financiamento de alguma agência de fomento para a realização de sua pesquisa, há acordos cujos termos estabelecem a devida citação da agência nas publicações geradas a partir do estudo. Não se trata apenas da referência de artigos resultantes de pesquisa, mas também dos dados obtidos durante o processo de produção científica. Por isso, a menção a agências de fomento deveria ocorrer pelo autor dos dados, bem como pelo pesquisador que optar por reutilizá-los em suas publicações. Convém destacar, ainda, que a disponibilização de dados está relacionada àqueles não confidenciais, os quais não podem ser expostos abertamente, sem nenhum crivo ético.

A anonimização de participantes da pesquisa pode ser entendida como o conjunto de procedimentos adotados por pesquisadores, com o propósito de inviabilizar a identificação de pessoas que participaram de um estudo científico. Tais procedimentos não se resumem ao uso de um nome fictício em citações textuais, mas incluem também, outras características que impedem que identidades sejam indevidamente reveladas, como o nome de lugares e instituições. Esse é um aspecto de relevância principalmente quando o pesquisador trabalha com temáticas sensíveis à sociedade, que dizem respeito, por exemplo, a questões criminais e execução penal. Portanto, algumas pesquisas da área requerem o cumprimento de termos de sigilo, nos quais dados não poderiam ser disponibilizados, publicamente, sem nenhum tipo de tratamento técnico.

Relacionado ao último aspecto mencionado, destaca-se que, em geral, o participante da pesquisa aceita os termos de participação com quem coordenou a pesquisa, e não, necessariamente, com quem reutilizou os dados disponíveis. Nesse sentido, dados poderiam ser expostos sem autorização de quem os concedeu e, por conseguinte, acordos de confidencialidade poderiam ser transgredidos.

O relato de PS1 sintetiza os aspectos até agora abordados em relação a tais

acordos:

Para nós produzirmos dados originais, nós precisamos de dinheiro, nós pedimos dinheiro para uma agência de fomento, nós prestamos conta a essa agência de fomento, e essa agência de fomento exige que o nome dela seja colocado nos artigos. Só que quando o dado é aberto, a pessoa que usa os dados não coloca o nome da agência de fomento nos artigos. Então, existe um dilema delicado aí, que isso não é nada resolvido. Um outro dilema é em relação a quem aceitou participar da pesquisa, quem aceitou participar da pesquisa aceitou mediante um acordo com quem coordena a pesquisa, quem obteve financiamento [...] vou dar um exemplo, assim: às vezes, a gente abre um dado e não tem o nome da pessoa lá, mas tem, por exemplo, a cidade, e a pessoa consegue identificar. Eventualmente, a pessoa consegue identificar a região, consegue, eventualmente, até identificar, por que por exclusão, ela consegue chegar em qual instituição foi feita a atividade, entende? Então, a pessoa que usa o dado, ela, por vezes, acaba ferindo acordos de confidencialidade, de anonimato.

Por outro ângulo, acordos éticos também deveriam ser observados para inibir a apropriação indevida de ideias ocasionada pela disponibilização de dados abertos e consequente perda do ineditismo de publicações. Isso porque pesquisadores ficam receosos de que outros cite procedimentos e dados da pesquisa, sem a devida atribuição de autoria. Ainda de acordo com PS1, na comunidade científica há pesquisadores que não prezam pelo trabalho colaborativo, mas sim, pelo trabalho competitivo. Esses, por sua vez, agem em prol da publicação antecipada de conteúdos científicos e, geralmente, não atendem a acordos éticos estabelecidos para a pesquisa, bem como explicou:

Então, você acaba prestando um serviço para alguém que se comporta, não como um colega de pesquisa, como um colega de academia. Ele se comporta como um competidor, se comporta como um alguém que quer publicar antes de você, porque ele não aceita participar do grupo de pesquisa e debater, coletivamente, os acordos para geração de dados de publicação. Ele quer pegar os dados e usá-los da maneira que ele quiser, sem respeitar nenhum tipo de acordo.

4.2.2.3 Categoria “Planejamento”

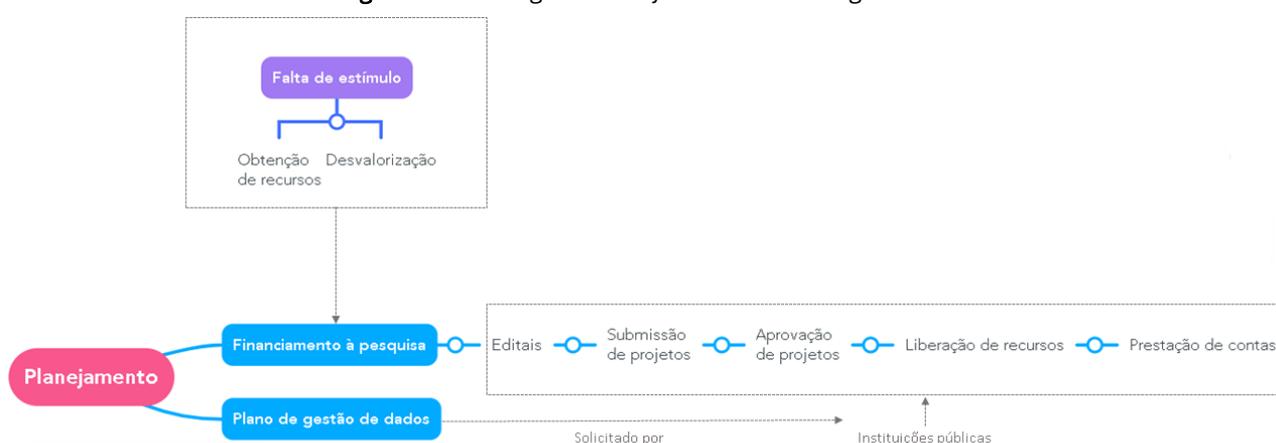
A categoria “Planejamento” está relacionada à organização dos processos envolvidos em pesquisas e ao gerenciamento de tais processos. Sendo assim, com base nos dados obtidos no estudo, essa categoria divide-se em duas subcategorias: “Financiamento à pesquisa” e “Plano de gestão de dados de pesquisa”.

Como representado na Figura 22, o financiamento à pesquisa refere-se aos trâmites necessários para a liberação de recursos que fomentarão estudos científicos. Geralmente, ocorre no âmbito de instituições públicas. Dessa maneira, são lançados editais de ampla concorrência, nos quais são submetidos projetos de pesquisa que serão avaliados por especialistas. Quando aprovados, tais projetos recebem recursos para serem executados.

Posteriormente, é comum haver a prestação de contas por parte do pesquisador que os recebeu. Toma-se nota, ainda, que pesquisadores se sentem desestimulados diante das dificuldades relativas à obtenção de recursos e à falta de valorização profissional, sendo essa última, acentuada pela ausência de políticas de reconhecimento do trabalho de pesquisadores no contexto brasileiro.

Já quanto ao Plano de gestão de dados de pesquisa, geralmente, esse documento é solicitado por instituições públicas responsáveis pelo financiamento de pesquisas.

Figura 22 – Categoria Planejamento - Sociologia



Fonte: elaborado pela autora

- Subcategoria “Financiamento à pesquisa”

Os dados obtidos nas entrevistas realizadas expressam o quanto o apoio financeiro de agências de fomento é essencial para o desenvolvimento de pesquisas na área de Sociologia, principalmente aquele oriundo de instituições públicas. Conforme relatos de PS3, PS4, PS5, PS6, PS7 e PS10, tais agências publicizam editais de ampla concorrência, nos quais são oferecidos recursos para o financiamento de pesquisas sobre temáticas específicas. Nessa direção, cientistas submetem seus projetos para avaliação das agências e, se aprovados, esses passam a ser integralmente custeados com verba destinada para sua execução.

Por isso, destaca-se que apesar de haver participação privada no apoio financeiro de pesquisadores entrevistados, a maior parte do financiamento decorre de instituições públicas, como explicitado por PS2, ao comentar que raramente obtém financiamento à

pesquisa por meio de iniciativas privadas. Aspecto semelhantemente importante foi abordado por PS6, no qual afirmou participar de editais de financiamento elaborados por instituições governamentais, como ministérios e secretarias. Trata-se, especificamente, de instituições indutoras de políticas públicas, conforme explicou:

Em geral, essa área, a gente tem demanda de pesquisa, por exemplo, por órgãos governamentais: ministérios, secretarias ou instituições policiais. Por exemplo, eu já pude fazer pesquisas assim, que são pesquisas propostas para alguma temática da gestão pública que eles queiram dar conta, e tem a minha agenda de pesquisa, vamos dizer 'pessoal e institucional' [...] no caso da primeira questão, que a gente pode chamar de pesquisas que são indutoras de políticas públicas, em geral, eles apresentam nas chamadas, nos editais quais seriam os objetos propostos, eu estou falando de um edital público, por exemplo, do Ministério da Justiça, do qual eu já participei algumas vezes e já ganhei [...] quando eles apresentam editais dessa natureza, eles já colocam, o gestor, né? O órgão demandante, já coloca a temática, o objeto, e a partir disso, eu vou construindo junto com a minha equipe de pesquisadores, tanto do núcleo, quanto de pessoas que estão sendo orientadas por mim, no mestrado e doutorado.

Com base nesse cenário de disponibilização de recursos é que pesquisadores estimam recursos necessários à pesquisa, de modo a otimizá-los. Por isso, tentam diminuir os custos da pesquisa, fazer parcerias com outras instituições e obter bolsas de iniciação científica para estudantes. Dessa maneira, o próprio desenho da pesquisa, a abordagem metodológica e os procedimentos analíticos são planejados para se enquadrarem nos recursos disponibilizados em editais. Ademais, pontua-se que os pesquisadores entrevistados destacaram o processo de prestação de contas sobre recursos recebidos por meio de financiamento público, no qual aplica-se um rigoroso método que demonstre com transparência a destinação apropriada dos mesmos.

Outro aspecto relevante apontado pela maioria dos entrevistados (PS1, PS2, PS3, PS6, PS8 e PS9) foi a falta de estímulo à pesquisa, decorrente principalmente da dificuldade na obtenção de recursos à pesquisa e da falta de uma política de reconhecimento do trabalho de pesquisadores no contexto brasileiro. Segundo PS1, não se trata apenas de recompensas financeiras, até porque muitas vezes um coordenador de pesquisa não recebe nenhum centavo a mais por liderar um estudo, mas é algo que diz respeito à valorização do coordenador, ao reconhecimento do conjunto de responsabilidades jurídicas, éticas e burocráticas que esse carrega em seus ombros. Por essa razão, ao invés da proposição de pesquisas originais que envolvem a produção de dados, seria mais produtivo a reutilização de dados públicos originários de órgãos oficiais, ou obtidos por meio de uma rede de pesquisa na qual o pesquisador não seja o coordenador do estudo. Isso porque os dilemas e responsabilidades envolvidos também seriam proporcionalmente menores.

A respeito das variações na disponibilidade de recursos por agências de fomento e das dificuldades do pesquisador em otimizá-los em decorrência de tais variações, PS3 comentou:

E vai montando [a pesquisa] de acordo com o recurso que é disponibilizado por esses editais. E aí depende, tem épocas que o CNPq vai disponibilizar editais de R\$ 200.000,00 e tem época que ele vai mal disponibilizar edital de R\$ 10.000,00. Então, a gente vai pensar no escopo da pesquisa a partir dessas dimensões [...] você fica o tempo todo buscando recurso e mais recursos. Como a gente está dentro de uma universidade, eu acho interessante também conversar com pessoas fora da universidade. Então, primeiro os nossos salários são dados, então, a gente não tem uma pressão tão grande de ter sempre recursos para pagar nossos salários, mas a gente acaba fazendo muito, submetendo a nossa proposta e edital para que a gente possa pagar as bolsas dos pesquisadores que trabalham com a gente, e aí são bolsas de pós-doc, mestrado, doutorado e graduação, mas dentro da nossa capacidade, que é limitada, já que somos uma equipe pequena.

Semelhantemente, PS2 ressaltou que vive uma maturidade científica adquirida ao longo dos anos em diferentes países, mas que se vê limitado com a falta de estrutura e condições para desenvolver suas pesquisas no Brasil. Com visão otimista, disse que continuará batalhando para obter recursos que serão aplicados diretamente nas pesquisas e que, apesar de ser um processo difícil, o que lhe move é a paixão pela ciência e a vontade de contribuir nessa direção.

De outro ângulo, pesquisadores como PS4 relacionaram o planejamento a trabalhos colaborativos realizados em redes de pesquisa. Geralmente, tais trabalhos também ocorrem mediante financiamento proposto em editais específicos e são instruídos por agências de fomento em relação à gestão de dados produzidos ao longo do estudo. Nesse contexto, a colaboração científica é vista como uma oportunidade na qual diversas equipes de pesquisa distribuídas em localidades distintas têm acesso ao financiamento de projetos.

- Subcategoria “Plano de gestão de dados”

Apenas quatro pesquisadores entrevistados afirmaram já ter experiência na elaboração de um Plano de gestão de dados. Em geral, pesquisadores desconheciam esse termo, no sentido da apresentação de um documento formal com esse título, ou outro semelhante. Entretanto, ficou nítido que eles lançam mão de diversos instrumentos que os auxiliam no planejamento da gestão de dados produzidos, principalmente, planilhas que os orientam quanto à própria organização da equipe de pesquisa. Como disse PS7, o termo “Plano de gestão de dados” seria muito representativo para abranger o tipo de planejamento que realiza em relação aos dados de suas pesquisas. Outros pesquisadores como PS5

afirmaram que nunca foram solicitados por qualquer instituição para elaborar um plano nesse sentido.

Por outro lado, alguns entrevistados afirmaram que já elaboraram um Plano de gestão de dados de pesquisa (PS1, PS3, PS4 e PS6). A partir dos relatos obtidos, pode-se observar que se trata de um documento trabalhoso, no qual pesquisadores apresentam dificuldades para a sua redação e sugerem mais apoio institucional para orientá-los nesse processo. Com base principalmente nos relatos de PS1, PS3 e PS6, percebeu-se que o Plano de gestão de dados deve ser um documento coeso que apresente com clareza os aspectos relativos aos dados para que os mesmos possam ser compreendidos e reutilizados por quaisquer pessoas.

De acordo com os pesquisadores mencionados, o Plano de gestão de dados é direcionado a alguma instituição solicitante e pode incluir informações relativas a estratégias utilizadas para a obtenção de dados, procedimentos de preservação de dados (tais como *backup*), descrição de acordos de confidencialidade e de consentimento livre e esclarecido estabelecido com participantes da pesquisa, apresentação de unidades a serem analisadas e dos procedimentos metodológicos que serão adotados no estudo. Também pode abranger a justificativa para a realização do trabalho científico, a descrição dos dados que compõem a pesquisa e até mesmo estar embasado na Lei de acesso à Informação (LAI), quanto à disponibilização de dados referentes ao estudo.

PS6 elaborou um Plano de gestão de dados, após solicitação da polícia civil de determinado estado, para a obtenção de bancos de dados relacionados a índices de criminalidade. Em razão disso, teve que redigir um documento robusto especificando todo o planejamento relacionado aos tipos de dados envolvidos no estudo, de modo a detalhar os parâmetros de coleta e sua destinação. O documento foi criteriosamente analisado por setores de gestão da informação e de análise criminal da instituição, até ser concedida a devida autorização para a liberação dos dados.

Como observado, o Plano de gestão de dados não se refere apenas à elaboração de um documento que visa cumprir exigências institucionais para a liberação de dados ou financiamento de pesquisas. Trata-se de uma descrição fidedigna sobre os processos relativos aos dados, desde a sua produção, até sua preservação a longo prazo. Por isso, pesquisadores como PS1 relataram que essa não é uma atividade trivial. Nesse sentido,

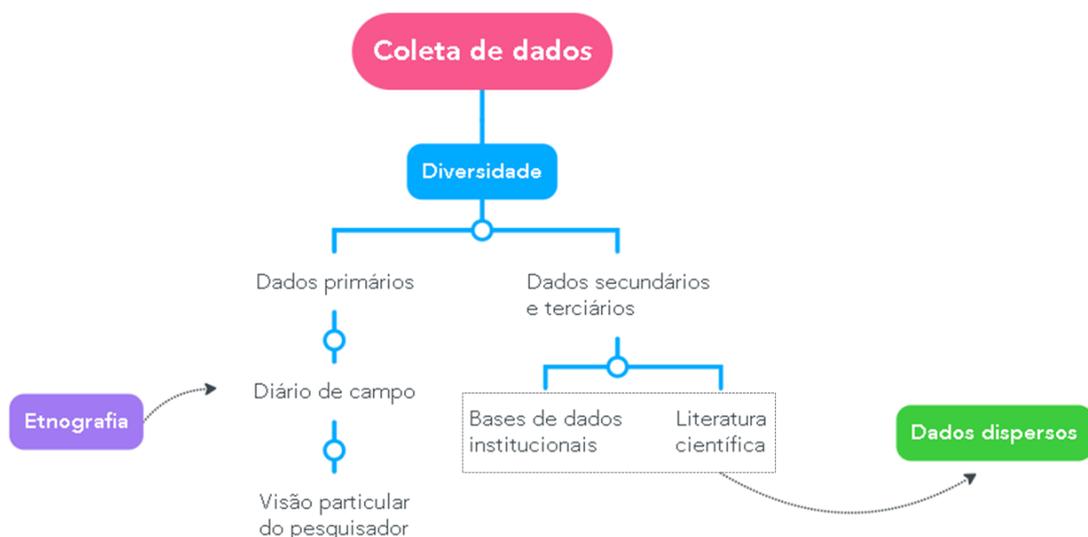
PS1 expôs essa dificuldade e mencionou que não há diretrizes brasileiras amplamente acessíveis que orientem o pesquisador quanto à elaboração de um Plano de gestão de dados:

[O Plano de gestão de dados] é um trabalho hercúleo, é muito trabalho [...] o problema é que ninguém ensina montar, você precisa entender. Então, em geral, você aprende olhando para o exterior, você olha em outros países como é que eles fazem, você tenta repetir os protocolos. E pensar que você está no Brasil, você tem que ter protocolos extras, só que é muito trabalho e quando você chega no final, depois dessa trabalhadeira toda, você entrega um papel para uma pessoa assinar.

4.2.2.4 Categoria “Coleta de dados”

A categoria “Coleta de dados” foi desenvolvida por meio da subcategoria “Diversidade”. Essa, por sua vez, representa o conjunto de naturezas e tipologias diferentes de dados de pesquisa. Trata-se de uma produção plural em que pesquisadores lançam mão de dados primários, como diários de campo registrados em meio a estudos etnográficos, e dados secundários e terciários, que abrangem dados dispersos em bases institucionais e na literatura científica. A Figura 23 representa essa categoria, em nível de suas relações:

Figura 23 – Categoria Coleta de dados – Sociologia



Fonte: elaborado pela autora

- Subcategoria “Diversidade de dados”

Com base nos dados obtidos no estudo, observou-se que os entrevistados produzem uma diversidade de dados, que é representada por meio de formatos distintos em manuscritos, material impresso ou digital. Sendo assim, dados são coletados mediante a aplicação de diferentes técnicas em contextos metodológicos qualitativos, quantitativos e mistos.

Nessa direção, pesquisadores relataram que produzem os dados nas pesquisas que realizam ou os reutilizam de outras fontes de informação. Na primeira situação, o pesquisador geralmente vai a campo e gera dados a partir de sua atuação na pesquisa com indivíduos, ambientes, objetos, entre outros. Desse modo, entrevistados como PS2 e PS7 mencionaram que é comum realizarem entrevistas e aplicarem questionários. Inclusive, PS2 destacou que, durante a pandemia da Covid-19, lançou mão de diferentes recursos tecnológicos para a coleta de dados, tais como aplicativos e correio eletrônico. Ademais, ressalta-se a importância de treinamento da equipe de coleta de dados quando ocorre a aplicação de questionário, na medida em que a padronização do procedimento pode minimizar a ocorrência de vieses em relação às respostas obtidas.

No que se refere à coleta de dados, pesquisadores como PS7 destacaram a utilização da etnografia como recurso metodológico em suas pesquisas. Nesse cenário, ressalta-se a importância do diário de campo como instrumento de registro de dados. Isso porque, na perspectiva dos entrevistados, tal material compreende a visão particular do pesquisador sobre os eventos que o cercam. Ou seja, trata-se do registro de seus relatos íntimos a respeito do fenômeno estudado.

Ainda sobre o diário de campo, PS4 fez o seguinte comentário:

São muitos os procedimentos, mas o básico da etnografia é a observação participante. Então, a gente observa, vive junto as rotinas com as pessoas que a gente trabalha, e depois a gente relata em diários de campo essa observação. Então, os diários de campo são o nosso material fundamental de pesquisa. Muitas vezes a gente faz também entrevistas e transcreve essas entrevistas, mas o básico da etnografia é a observação. Por exemplo, se eu tivesse fazendo etnografia aqui, eu terminaria a nossa conversa e eu relataria em um diário o que aconteceu nessa conversa aqui e nessa interação que a gente está tendo. Como que ela foi pensada, como que ela aconteceu, o que a gente conversou, como que você é, que a gente descreve as pessoas, os personagens, etc., os lugares e tudo isso. E guardaria esses dados comigo, guardaria esse diário de campo comigo, como eu tenho muito diário de campo guardado, esse é o fundamental.

Entrevistados também afirmaram reutilizar dados originários de outras fontes de informação, tais como instituições governamentais. A título de ilustração, PS2 disse que

utiliza dados censitários que incluem amostra de domicílio, censo demográfico e censo agropecuário. Semelhantemente, PS8 comentou que acessa bases de dados da área da saúde e de indicadores socioeconômicos, como o departamento de informática do Sistema Único de Saúde do Brasil (DATASUS) e o Instituto Brasileiro de Geografia e Estatística (IBGE), respectivamente, em razão da diminuição de custos para a obtenção de dados de pesquisa.

Entretanto, observou-se que muitos desses dados institucionais encontram-se dispersos, ou seja, as instituições alimentam suas bases, e para acessá-las, o pesquisador deve localizar cada uma delas e realizar buscas em diferentes plataformas. Portanto, na perspectiva de pesquisadores como PS6, as informações estão descentralizadas e demandam ser organizadas e disponibilizadas em uma única plataforma interoperável, conforme exemplificou:

Para eu conseguir acessar um determinado dado, eu tenho que ir a cada instituição. Já aconteceram esforços pelo governo brasileiro, pelos estados e municípios de compatibilizar bases de dados, o que seria muito mais fácil para gente, mas isso ainda não é uma realidade [...] a gente carece muito de um repositório adequado, sabe? [...] então, você encontra um pedaço, você não encontra outro. A gente vive em zonas cinzentas, eu imagino que em outras políticas públicas também esse fenômeno acontece, mas no nosso, que a gente está lidando com a segurança das pessoas, a vida das pessoas, isso é muito complicado, para dizer o mínimo, né?

Além disso, entrevistados também comentaram que reutilizam dados provenientes da própria literatura científica, principalmente, para construir a base teórica de revisões integrativas. Esse tipo de estudo é fundamental para a caracterização do estado da arte de determinada área do conhecimento.

4.2.2.5 Categoria “Descrição de dados”

Tal como é apresentada na Figura 24, a categoria “Descrição de dados” foi desenvolvida por intermédio da subcategoria “Padrão local”. Por seu turno, essa refere-se a padrões de descrição de dados estabelecidos pelos próprios pesquisadores que os manuseiam, como a adoção de categorias no registro em diários de campo e preenchimento de campos descritivos em planilhas eletrônicas. Nesse contexto, destaca-se que nenhum entrevistado no estudo relatou adotar padrões internacionais de metadados para a descrição de dados, ainda que já existam padrões relacionados à área de Ciências sociais e Humanidades, tais como o DDI e o SDMX.

Figura 24 – Categoria Descrição de dados - Sociologia



Fonte: elaborado pela autora

- Subcategoria “Padrão local”

Os dados obtidos nas entrevistas demonstraram que a descrição de dados de pesquisa é realizada pelos próprios pesquisadores, que, por sua vez, adotam o seu padrão para melhor organizá-los e recuperá-los. Ou seja, ainda não utilizam metadados internacionalmente padronizados para esse fim.

A título de exemplo, PS4 comentou sobre a descrição de conjuntos de dados registrados em diários de campo. Assim, PS4 estabeleceu cinco categorias nas quais os dados geralmente são descritos: tempo (aspectos cronológicos), espaço (dimensões cartográficas), descrição densa de situações (detalhamento das interações ocorridas na pesquisa), *insights* analíticos (percepções analíticas do pesquisador) e diário íntimo (ponderações pessoais do pesquisador no campo de estudo).

PS9 comentou que os dados obtidos são caracterizados por descritores básicos, como: título, autor, resumo e palavras-chave. PS10 afirmou que não utiliza padrões de metadados internacionais da área, mas realiza uma descrição local de dados por meio do preenchimento de campos de planilhas no *Excel*. PS6 também descreve dados de suas pesquisas de modo particular. Por meio de planilhas eletrônicas, costuma atribuir metadados que incluem informações sobre localidade, gênero, etnia, profissão e evento (variável geralmente associada à tipificação penal). A respeito disso PS6 explicou:

A gente disponibiliza as planilhas em Excel descrevendo as variáveis que a gente está trabalhando. Então, idade, eu vou dar um exemplo do que eu faço que aí fica mais fácil para mim. Então, sei lá, idade, gênero, raça, local de moradia, profissão. O evento, o evento de violência e criminalidade, em geral o evento tem a ver com o que está descrito no código penal. Então, foi um homicídio, foi um latrocínio, então, esse é um descritivo importante porque aí outro pesquisador pode cruzar de novo esses dados. A gente vai colocando, disponibilizando a base

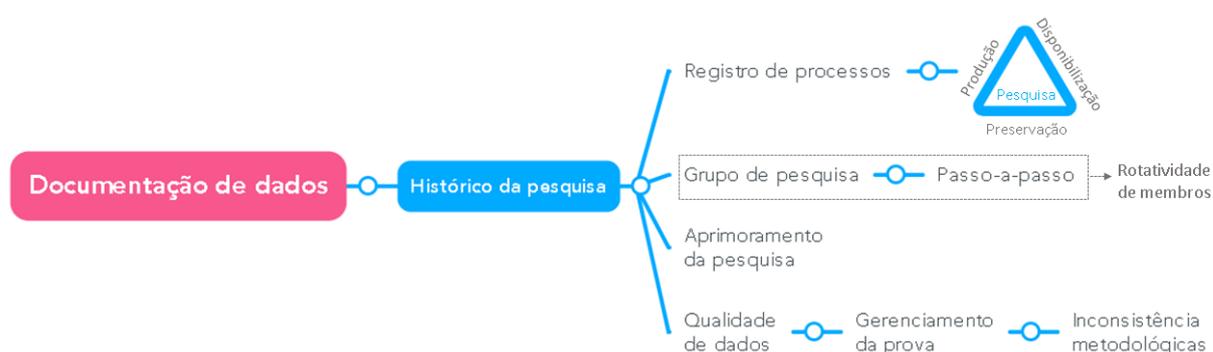
com esses descritores, vamos dizer assim, relacionados àquele conjunto de informações que a gente conseguiu de algum órgão. Enfim, é mais ou menos isso.

Apesar de nenhum pesquisador ter mencionado a adoção de padrões internacionais de metadados amplamente utilizados, já é conhecida a existência de padrões especificamente destinados à área de Ciências sociais e Humanidades, como o *Data Documentation Initiative* e o *Statistical Data and Metadata Exchange (SDMX)*. Um outro padrão de metadados comumente utilizado, porém de natureza generalista e adotado em diferentes áreas, é o *Dublin Core*.

4.2.2.6 Categoria “Documentação de dados”

A categoria “Documentação de dados” foi expandida em termos da subcategoria “Histórico da pesquisa”, que abrange quatro aspectos principais: 1) registro de processos da pesquisa, desde a produção, até a disponibilização e preservação de dados; 2) descrição do passo-a-passo de estudos no âmbito de grupos de pesquisa; 3) aprimoramento de estudos realizados; 4) ateste da qualidade de dados, por meio do gerenciamento da prova e verificação de inconsistências metodológicas. A Figura 25 representa essa categoria e suas relações:

Figura 25 – Categoria Documentação de dados - Sociologia



Fonte: elaborado pela autora

- Subcategoria “Histórico de pesquisa”

Conforme observado nas entrevistas, principalmente as concedidas por PS1, PS3, PS6, PS9 e PS10, a documentação sobre os dados de pesquisa é muito importante para

registrar os procedimentos relativos aos dados, desde a sua produção até sua disponibilização e preservação a longo prazo. Mas sua relevância não se restringe a tal aspecto. A documentação também é fundamental para ampliar a compreensão de pesquisas realizadas por meio de seu histórico, ou seja, da descrição dos caminhos trilhados durante os estudos.

Estar a par do histórico de pesquisas é bastante útil em um grupo de pesquisa em que a rotatividade de membros é um fato. Isso porque muitos desses grupos são integrados por estudantes de mestrado, doutorado e bolsistas de iniciação científica na graduação, que em geral permanecem na equipe até a duração do curso de pós-graduação e de graduação, respectivamente.

Portanto, o registro do passo-a-passo em relação aos dados amplia a compreensão da pesquisa por parte de novos membros da equipe, bem como armazena os processos realizados por pesquisadores que já não integram o grupo. Nessa direção, PS1 comentou:

O principal é que as pessoas saem ao longo do processo de pesquisa e os cabelos brancos atestam o envelhecimento, você esquece, você tem que documentar porque as pessoas simplesmente saem e você perde aquele histórico, perde aquele conhecimento, você perde. Então, você tem que documentar porque você não pode depender das pessoas. Então, todo procedimento precisa ser documentado.

Ao falar sobre documentação de processos relativos a dados, PS10 relatou que o primeiro nível desse processo é referente à aprovação do projeto de pesquisa de seus orientandos junto ao comitê de ética, sendo necessária a guarda de todos os documentos associados ao trâmite. Outro aspecto apontado por PS10 foi o registro do andamento de pesquisas no *Google drive* do laboratório que coordena. Tal medida é fundamental para que todos os integrantes do grupo saibam quais dados estão sendo produzidos e sobre o que abordam. Nessa direção, PS10 comentou que mesmo que não haja nenhum interesse em reutilizar dados disponíveis na plataforma em determinado momento, esses continuam acessíveis para qualquer estudo futuro. Ademais, PS10 enfatizou que a rotatividade com mestrandos e doutorandos é muito grande, todavia, os dados produzidos no laboratório, ali ficam registrados.

O histórico dos processos relativos aos dados de uma pesquisa também possibilita o desenvolvimento de novos estudos, que inclusive, podem ampliar o escopo e aprimorar a pesquisa inicial. Por isso, PS6 disse que a falta desse detalhamento constitui óbice para o desenvolvimento da área, como se pode observar:

Nossa! É fundamental! A gente tem essa prática. Assim, não só porque é uma questão do mundo científico de você ter uma parte dedicada a metodologia, as suas estratégias metodológicas, mas a gente tem reforçado muito isso, porque a ideia é que outras pessoas possam produzir pesquisas que complementem, que aprofundem coisas que a gente não conseguiu avançar e que façam contrapostas àquilo que a gente está propondo. Então, se você não descreve e não detalha esses processos, fica muito, muito complicado. Então, a gente tem se dedicado muito mesmo, inclusive, falando desses óbices, né?

Outro aspecto relevante apontado por PS3 foi que o histórico de pesquisas contido na documentação de dados também é útil para atestar a qualidade de dados. Em determinadas situações, pesquisadores têm uma imagem enviesada de que dados representam além do que de fato eles contém. Por isso, é fundamental recorrer à documentação relativa aos mesmos para a averiguação de informações, relacionadas, por exemplo, à representatividade dos dados quanto à coleta.

Associada à questão da qualidade, está a comprovação de dados por meio de documentação, conforme mencionaram PS5, PS7 e PS8. Sobre esse ponto, PS5 explicou o que chamou de “gerenciamento da prova” para demonstrar a importância de se documentar as provas de que um estudo foi realizado. Tais provas encontram-se materializadas em registros de entrevistas, questionários, documentos ou outros instrumentos que indiciam que fenômenos foram investigados. Ao falar de suas próprias experiências relacionadas à documentação de dados, PS5 comentou:

Por exemplo, eu tenho um [estudo] com 70 questionários, isso que dá umas 20 entrevistas, está tudo guardado aqui, né? Os questionários, eu digitalizei na época. Alguns vieram impressos, outros vieram pela Internet. Os impressos eu digitalizei, guardei. Isso é a demonstração, isso é a prova, né? Isso é na sociologia, a gente chama o gerenciamento da prova. Então, o gerenciamento da prova é geralmente a entrevista, documentos, questionários, né? [...] A gente guarda porque faz parte mesmo do procedimento científico, guardar as provas dos dados que já tem.

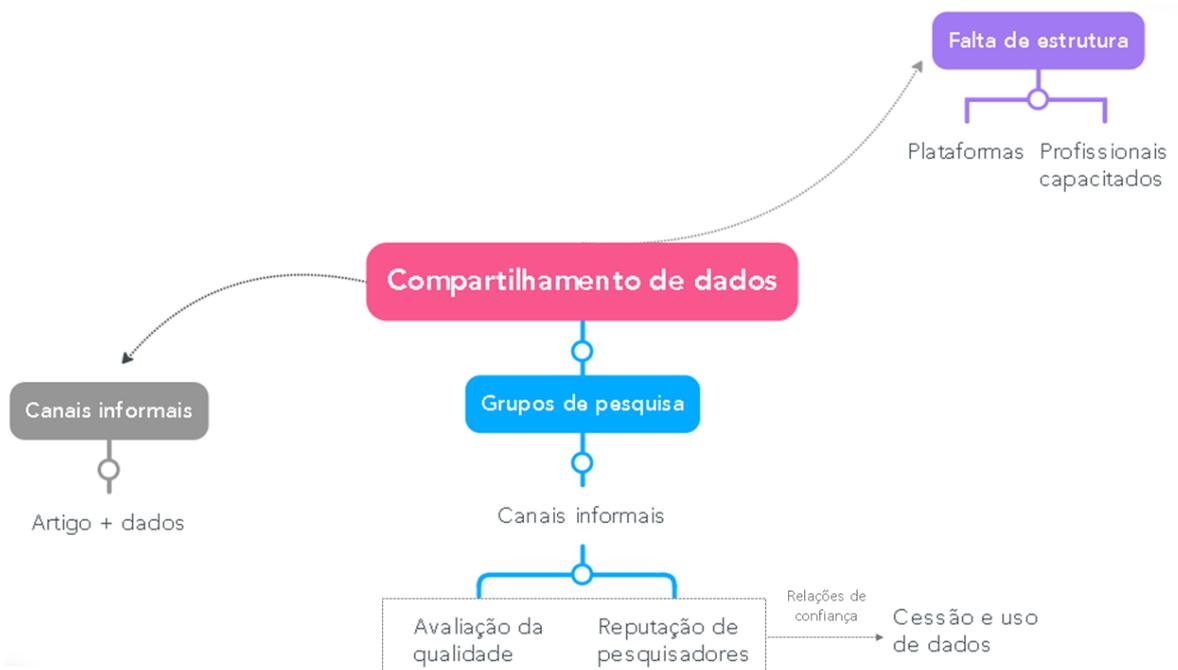
Nessa direção, PS8 reiterou a possibilidade de comprovação de dados a partir da documentação de pesquisas, de modo a facilitar a identificação de inconsistências metodológicas. De maneira semelhante, PS7 disse que esse procedimento viabiliza o ateste da veracidade de dados, quando os mesmos estão sendo avaliados. Mesmo quando não está sendo auditado por alguma instituição, PS7 se preocupa em documentar processos relativos às pesquisas que realiza, conforme ficou perceptível em sua fala:

Porque assim, as agências de pesquisa não vão me cobrar isso: ‘ah você fez tal entrevista, ou não fez?’ ‘Me mostra aqui a transcrição!’ Não é o caso, mas eu acho importante do ponto de vista de: ‘ah não! Mas assim, esse levantamento foi feito?’ Eu tenho lá, se quiserem em algum momento existir uma auditoria em relação à pesquisa pública, que eu acho muito difícil que isso ocorra, e se ocorrer, a gente vai estar em um estado de exceção, eu tenho o material, eu tenho sabe? É mais no sentido de comprovar mesmo a pesquisa.

4.2.2.7 Categoria “Compartilhamento de dados”

Conforme mostrado na Figura 26, a categoria “Compartilhamento de dados” foi desenvolvida por meio da subcategoria “Grupos de pesquisa”. As informações obtidas no estudo demonstraram que dados são compartilhados por intermédio de canais formais, como periódicos científicos. Porém, verificou-se que tal compartilhamento ocorre, predominantemente, no âmbito de grupos de pesquisa, a partir de canais informais. Isso porque há relações de confiança nas quais se avaliam a qualidade de dados e a reputação de pesquisadores para a devida cessão de dados. Ademais, destaca-se que o compartilhamento de dados é influenciado pela falta de estrutura decorrente, principalmente, da indisponibilidade de plataformas e profissionais capacitados para lidar com dados.

Figura 26 – Categoria Compartilhamento de dados – Sociologia



Fonte: elaborado pela autora

- Subcategoria “Grupos de pesquisa”

Com base nos dados obtidos nas entrevistas, observou-se que alguns pesquisadores já tiveram experiências no compartilhamento de dados em canais formais, como os

periódicos científicos. PS5 comentou a respeito da decisão de tornar públicos e acessíveis os dados de uma de suas pesquisas, juntamente com o artigo submetido a uma revista. Nessa ocasião, PS5 ressaltou que tal revista não condicionou a publicação do artigo ao envio dos dados de pesquisa, porém, o periódico demonstrou algumas vantagens que estimularam sua escolha em disponibilizar os dados, como o alcance de maior visibilidade da pesquisa em nível internacional.

Outros pesquisadores, como PS6, relataram que ainda não possuem experiência na disponibilização de dados abertos, todavia, demonstraram grande interesse em assim fazê-la por meio de uma plataforma ou repositório de acesso público. Nessa direção, PS6 chegou a solicitar uma verba para a construção de um portal que disponibilizasse os dados de determinado projeto de pesquisa, mas que ainda não foi implementado. Sobre esse aspecto, PS6 comentou:

Eu pedi uma verba para construir esse portal para disponibilizar todos os dados do projeto [nome do projeto de pesquisa]. Era um pouco essa ideia, era uma verbinha curta. Eu me inspirei em uma colega, que é minha parceira de pesquisa que já faz isso. Ela tem sempre essa preocupação de no final do projeto sempre ter um blog, é um site mesmo, sabe? Outras pessoas podem baixar as coisas e tal, então, é isso.

Apesar dos relatos mencionados, sobre compartilhamento de dados em canais formais de publicação, tais como periódicos científicos e repositórios institucionais, a pesquisa verificou, a partir dos dados obtidos, que a maioria dos pesquisadores (PS1, PS2, PS4, PS8, PS9 e PS10) compartilham dados de modo informal, com colegas de departamento e, principalmente, nos grupos de pesquisa os quais integram.

PS2 destacou algo bastante interessante relacionado ao compartilhamento de dados entre pesquisadores conhecidos, que extrapola a avaliação da qualidade de determinado dado. Trata-se de uma prática influenciada por aspectos pessoais, na qual pesquisadores se conhecem pessoalmente e mantém relações de confiança, tanto para a cessão, quanto para o uso de dados. Dessa maneira, PS2 mencionou o aspecto da reputação do indivíduo o qual se pretende compartilhar dados, como se pode observar em uma de suas falas:

Eu acho que vai muito do plano pessoal, um pouco da confiança entre as pessoas, porque vale muito a reputação. Tu constroi uma reputação não é só em cima dos seus artigos, é em cima da pessoa que tu és. E aí, eu acho que as pessoas: 'ah fulano' é aquilo, sabe? Ah! Aquele lá, já trabalhei com ele, é alguém que eu posso compartilhar! Eu acho que entra muito para esse lado.

Entrevistados como PS4 apontaram algo comum na área da Sociologia, que é o registro de dados em diários de pesquisa. Como já mencionado, tal instrumento revela a

visão particular do pesquisador sobre os fenômenos estudados, trazendo consigo a sua autorreflexão obtida no trabalho de campo. Pelos motivos expostos, PS4 ressaltou que o compartilha apenas com colegas próximos ou integrantes do grupo de pesquisa o qual coordena. Todavia, esse material recebe um tratamento de anonimização antes de ser compartilhado. Além disso, é necessário que o mesmo receba uma ordenação sequencial para que a narração dos fenômenos observados seja lógica, não prejudicando assim, a interpretação dos registros. Nesse sentido PS4 explicou:

Eu já tive a experiência de disponibilizar, sim, mas sempre para colegas e participantes da nossa equipe. A minha pesquisa individual, os meus dados, os diários de campo, etc., foram, já muitas vezes, disponibilizados internamente aos projetos de pesquisa, mas eu não tenho a experiência de disponibilizá-los abertamente [...] porque o diário, em geral, é para ser lido pelos próprios pesquisadores. A gente coloca o nome das pessoas, coloca o nome dos lugares, então, para publicar você tem que mexer em tudo isso. Além disso, como os diários são feitos em sequência, muitas vezes, o leitor que lê um diário só, ele não vai pegar as referências. É que nem ver um capítulo de uma novela que você nunca viu, você vai ter um tempo para se adaptar ao que está acontecendo ali, quem são aquelas pessoas, qual que é a história delas e tal.

Nota-se também, que o compartilhamento de dados é uma prática predominantemente individual e pessoal, ainda não institucionalizada no ambiente dos pesquisadores entrevistados, conforme enfatizou PS10 e PS1. Esse último, comentou que tal prática ainda não é amplamente realizada, em razão da insuficiência de infraestrutura tecnológica e de recursos financeiros destinados a esse fim. Sendo assim, PS1 lembrou que certa vez foi solicitado por uma instituição pública a disponibilizar dados de uma pesquisa. Isso devido a um requerimento baseado em dispositivos contidos na Lei de acesso à informação. Todavia, enfatizou que todos os procedimentos foram realizados em nível institucional, e não individual.

Caso tivesse que disponibilizar abertamente dados, hoje, PS1 disse que não poderia fazê-lo porque não possui estrutura, tais como uma plataforma destinada a esse fim, ou profissionais capacitados para dar-lhe instruções sobre como proceder a esse respeito. Sendo assim, PS1 fez o seguinte comentário quando foi perguntado sobre sua experiência no compartilhamento de dados abertos de pesquisa:

PS1: Olha, até já, mas não fui eu que disponibilizei individualmente. Foi disponibilizado pela instituição que estava fomentando a pesquisa. Entrevistadora: Foi uma exigência da instituição? PS1: Não, era um órgão federal e alguém solicitou transparência, solicitou o acesso aos dados pela lei da transparência, alguma coisa assim, é que me lembro. Isso tem mais de uns 15 anos, assim. Ah, o dado foi disponibilizado e tudo bem, não deu nenhum problema não! [...] então, respondendo a sua pergunta, já houve sim, mas não o fiz no meu CPF, foi algo feito pelo órgão que financiava e tudo mais e eu, hoje, não disponibilizaria, a não ser que coisas muito pequenas e tal, porque eu não tenho... eu não tenho nem condições de fazê-lo, de ter um servidor que torne isso público. Você tem que ter um servidor, tem que ter um site, você tem que ter alguém que alimente esse

site. *Eu não tenho esse dinheiro, eu não tenho condição de fazê-lo e hoje eu teria muito receio de fazê-lo.*

4.2.2.8 Categoria “Preservação de dados”

Assim como é representado na Figura 27, a categoria “Preservação de dados” foi desenvolvida em termos da subcategoria “Estratégias locais”. No que lhe toca, tais estratégias de preservação são realizadas pelos próprios pesquisadores. Caracterizam-se como sendo não institucionalizadas, uma vez que são individuais, além de variarem de acordo com as disciplinas em que são adotadas. Nesse contexto, é notório que há riscos de perdas de dados.

Figura 27 – Categoria Preservação de dados – Sociologia



Fonte: elaborado pela autora

- Subcategoria “Estratégias locais”

Nenhum dos pesquisadores entrevistados afirmou que já elaborou algum Plano de preservação de dados. Inclusive, esse é um documento pouco conhecido dentre os mesmos. Com base em todos os relatos obtidos, foi possível observar que a preservação de dados ocorre mediante a utilização de estratégias dos próprios pesquisadores. Em geral, essa prática não é institucionalizada nas universidades em que atuam, ou seja, trata-se de medidas individuais e que seguem os padrões de guarda e cópias de segurança do próprio laboratório.

Em síntese, podem ser elencadas as seguintes estratégias de preservação de dados apontadas pelos entrevistados:

- armazenamento de dados em computadores do laboratório de pesquisa de universidades;
- armazenamento de dados em *Hard Disk* (HD) e mídias físicas, como *Digital Versatile Disc* (DVD) e *Compact Disc* (CD);
- armazenamento de dados em nuvem;
- guarda de documentos físicos em arquivos, tais como questionários e entrevistas;
- recomendações de pesquisadores a orientandos sobre a importância de guarda e preservação de dados de pesquisa.

Ao falar sobre preservação de dados, PS8 destacou que no campo da Sociologia essa é uma atividade comumente atribuída ao próprio pesquisador. PS8 ainda sugeriu que a preservação em outras áreas, como a Química e as Ciências médicas, deva ocorrer em caráter institucional, com a proposição de níveis de acesso a dados para garantir a segurança nesse contexto. Sobre isso PS8 explicou:

No caso das pesquisas que eu realizei na Sociologia, essa preservação dos dados era com a gente mesmo, você deixava em um HD, em uma gaveta e pronto. Você não tinha nenhum, sei lá como posso dizer, esses níveis de segurança onde isso vai ficar armazenado, enfim, não. Aqui na Sociologia, não. Acredito que em outros departamentos da [universidade em que atua] sim, principalmente, talvez na área química deva ter também onde você tem essa questão de laboratórios, enfim, dados médicos da saúde, provavelmente deva ter algum espaço onde isso é armazenado, onde você tem essa questão de material biológico, ou onde alguma pesquisa pode gerar, por exemplo, patente. O departamento de química aqui produz muita patente, então, isso deve ter uma segurança maior para garantir, mas, no caso da Sociologia é aquilo, você põe no HD e guarda e pronto, não teria nenhum procedimento formalizado, algum lugar onde você colocaria esses dados, níveis de segurança, de acesso, enfim, não.

Como se pode observar, os dados produzidos pelos entrevistados variam quanto à natureza e formatos. Dessa maneira, incluem entrevistas, questionários e transcrições em meio físico ou digital, vídeos, áudios, entre outros. É conveniente que cada tipo de dado seja armazenado conforme às suas especificidades e formatos, o que parece não ter sido tarefa fácil aos pesquisadores. Outro aspecto interessante, é que nenhum pesquisador afirmou depositar dados produzidos em um repositório de dados gerido pelas universidades as quais atuam, apesar de alguns deles expressarem a vontade de assim fazê-lo.

Conforme denominou PS10, a preservação de dados no âmbito das Ciências sociais soa como um conjunto de procedimentos artesanais, ocorridos em nível individual. Consequentemente, fica nítido que tal situação expõe dados de pesquisa a riscos de perdas devido à obsolescência de mídias e, principalmente, prejuízos causados por danos naturais,

que podem ocasionar perda física de servidores computacionais e também de materiais impressos.

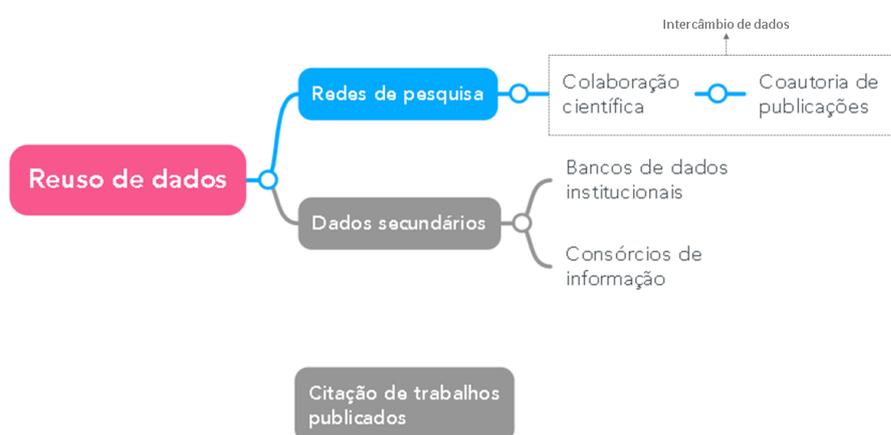
Em seu relato, PS9 ilustrou que ao longo de sua carreira, utilizou diversas mídias com a finalidade de armazenamento de dados. Entretanto, muitos deles se perderam ou já não podem mais ser acessados, em razão da defasagem de aparelhos incapazes de fazer a leitura adequada dos mesmos. Nessa direção, PS9 comentou:

Esse problema de trabalhar com mídias diferentes, então, há 20 anos a gente tem lá um [nome do núcleo de estudo da universidade], que é o núcleo ligado às pesquisas que eu faço. Então, tem esse material de entrevistas, por exemplo, em áudios e também transcritas e digitalizadas, né? Que estão gravadas nesse tipo de mídia, né? Ou seja, na verdade são discos e, às vezes, a gente vai perdendo, né? Quer dizer, por exemplo, tem disco lá que daqui a pouco vou ter que usar o aplicativo porque a interface já não bate. São esses problemas que os museus enfrentam, que os arquivos enfrentam, e a gente não tem muita, não tem capacidade técnica de lidar com isso.

4.2.2.9 Categoria “Reuso de dados”

Conforme demonstra a Figura 28, a categoria “Reuso de dados” é compreendida por intermédio da subcategoria “Redes de pesquisa”. Os dados obtidos no estudo mostraram que pesquisadores reutilizam dados secundários, sobretudo, oriundos de bancos de dados institucionais e consórcios de informação na área de Ciências Sociais. Além disso, reutilizam trabalhos já publicados, citando-os por meio da elaboração de referências bibliográficas. Entretanto, verificou-se neste estudo que tal prática concentra-se no contexto de redes de pesquisa, por meio de colaboração científica e consequente coautoria em publicações científicas.

Figura 28 – Categoria Reuso de dados – Sociologia



Fonte: elaborado pela autora

- Subcategoria “Redes de pesquisa”

Com base nos relatos obtidos, principalmente o de PS2, PS5 e PS10, tomou-se nota de que o reuso de dados ainda não é uma atividade habitual dentre os entrevistados. O mais comum é que esses façam citação de trabalhos já publicados, e não dos dados, propriamente ditos. As falas de PS10 ilustram esse aspecto, na medida em que se afirma que o reuso de dados não é comum na área e que, geralmente, se utiliza resultados publicados por colegas de trabalho, não os dados primários:

[O reuso de dados] não é comum não, mas claro, a gente utiliza, mas não o dado primário, utiliza os estudos feitos, os resultados publicados por colegas de trabalho sobre o mesmo tema, ou do nosso grupo de pesquisa. Mas assim, reutilizar a entrevista feita, reutilizar o questionário: ‘ah, tal pesquisa, nós aplicamos um questionário e depois só conseguimos dar o tratamento em uma parte dele, tem aspecto que tenha ficado de fora’, não! Até se comenta isso: ‘ah, o nosso questionário tem muita, nós acabamos tendo muito mais informação do que a gente conseguiu analisar, ah, então quem sabe, quem sabe em outro momento a gente volte nesse material’.

No que se refere ao reuso de dados secundários, PS6 e PS9 mencionaram a utilização de bancos de dados institucionais e de plataformas de consórcios de informação na área de Ciências sociais. Sendo assim, destaca-se o acesso à banco de dados disponibilizados pelas seguintes entidades: Fundação Getúlio Vargas (FGV), Instituto de Pesquisa Econômica Aplicada (Ipea), Fundo das Nações Unidas para a Infância (Unicef) e Organização das Nações Unidas para a Educação, a Ciência e a Cultura (Unesco). Sob tal perspectiva, trata-se de dados que abordam substancialmente aspectos socioeconômicos, étnicos e culturais.

Apesar do exposto, o estudo verificou que quando há o reuso de dados, essa prática ocorre, predominantemente, no contexto de redes de pesquisa. Por sua vez, tais redes caracterizam-se, sobretudo, pela realização de trabalhos colaborativos e publicação de trabalhos em coautoria. Conforme declarado por PS1, PS4 e PS5, o reuso de dados é preponderante nos grupos de pesquisa os quais participam, sendo comum também o intercâmbio de dados produzidos por seus orientandos.

De modo semelhante, PS7, explicou que o reuso de dados no âmbito de pesquisas colaborativas com seus orientandos é bastante comum. Além disso, destacou que não costuma reutilizar dados de pesquisas os quais não foram supervisionadas por ele, como se pode notar em seu comentário:

Não, não, no interior do meu grupo, sim, né? Por exemplo, tem eu e mais dois que estamos trabalhando, dois orientandos de doutorados, estamos trabalhando com o mesmo objeto de pesquisa. Então, por exemplo, a gente está fazendo entrevistas com [participantes da pesquisa]. Então, a gente conversa e olha: ‘eu tenho essa

entrevista aqui, pode te interessar!’ Mas é mais no sentido de confrontar informações para ver se a gente está no mesmo caminho. Mas eu acho que nunca ocorreu isso, de a gente usar o dado de uma outra pesquisa em um texto individual, entendeu? Quando eu estou escrevendo uma coisa, eu não utilizo o dado de uma outra pesquisa que não foi feita por mim, ou sobre a minha supervisão.

À vista do exposto, convém ressaltar aspecto mencionado por PS3 a respeito do reuso de dados de pesquisa. Segundo PS3, essa é uma prática ainda pouca disseminada no Brasil, principalmente, porque dados se encontram concentrados nas mãos daqueles que os produziram. Por conseguinte, outros pesquisadores não têm amplo acesso aos mesmos, dificultando assim, a ação de pensar em novas questões a partir de dados já existentes. A título de ilustração sobre quais medidas podem ser tomadas em relação à disseminação da prática de reuso de dados, PS9 sugere a utilização de bancos de dados disponibilizados por consórcios de informações na área de Ciências Sociais.

4.2.2.10 Categoria “Gestão de dados abertos de pesquisa”

Assim como é representado na Figura 29, a categoria “Gestão de dados abertos em universidades” é entendida em termos de duas subcategorias: “Debates sobre o tema” e “Diferenças disciplinares”.

A primeira refere-se à importância da promoção de discussões a respeito dos dados de pesquisa no ambiente de universidades, como por exemplo, a realização de seminários. Nesse contexto, é fundamental a interação com a comunidade acadêmica, de modo a oportunizar a construção coletiva de conhecimentos sobre o tema. Até porque, pesquisadores entrevistados comentaram a falta de diretrizes sobre a gestão de dados de pesquisa e a sobrecarga de trabalho a qual tem sido expostos.

Por seu turno, a segunda subcategoria – “Diferenças disciplinares” – abrange as diversas perspectivas das áreas do conhecimento relacionadas à gestão de dados de pesquisa. Tais diferenças são importantes nesse contexto e sugerem a necessidade da realização de debates a respeito do tema, sob enfoque de construção coletiva. Sendo assim, a Figura 29 apresenta a referida categoria e as relações estabelecidas com suas subcategorias.

Figura 29 – Categoria Gestão de dados abertos em universidades - Sociologia



Fonte: elaborado pela autora

- Subcategoria “Debates sobre o tema”

Os dados obtidos no estudo, principalmente decorrentes das entrevistas com PS6, PS7, PS8 e PS9 apontam a necessidade de intensificar debates sobre a gestão de dados no contexto de universidades, principalmente em relação a departamentos e programas de pós-graduação. Algumas das maneiras pelas quais isso poderia ocorrer seria a realização de seminários, *workshops*, capacitações, debates e a ministração de disciplinas sobre o tema na modalidade livre. Nessa direção, PS6 afirmou ainda não ter visto nenhuma discussão sobre essa temática envolvendo os programas de pós-graduação da universidade em que atua, conforme o seu relato:

Eu acho que a gente deveria investir é em seminários internos sobre esse tema, porque provocada pela sua pesquisa eu comecei a querer. Quando eu te pedi o roteiro, eu fiquei lendo umas coisas tão interessantes sobre o seu tema, que eu não vi ainda uma discussão na [nome da universidade em que atua] envolvendo os programas de pós-graduação.

Semelhantemente, PS7 destacou a importância de a gestão de dados abertos de pesquisa ser amplamente discutida em universidades. Entretanto, deve-se pontuar que para PS7, isso deve ser resultado de uma construção coletiva entre a comunidade acadêmica, e não apenas uma decisão hierárquica que deve ser cumprida. Nesse sentido, PS8 também mencionou que a referida temática não tem sido pautada em discussões na universidade em que atua. Sendo assim, sugeriu que houvesse uma instituição centralizadora que fosse incumbida de disponibilizar os dados produzidos nas universidades e que elaborasse, com

clareza, a política de acesso adotada. A fala de PS8 reflete os aspectos abordados em relação a gestão de dados abertos de pesquisa:

Mas no caso da [nome da universidade em que atua] acho que é algo que precisa ser discutido, eu nem sei se seria no âmbito da universidade, dado o custo que isso deve gerar. Por exemplo, se houvesse algo no CNPq, uma parte lá que é disponibilização de dados de pesquisa. Enfim, alguma coisa centralizada, onde todos poderiam acessar. De repente, em termos de pensar uma política de acesso a essa produção de dados, se fosse talvez em algum ponto onde as pessoas reconhecessem e fossem lá.

Entretanto, observou-se, sobretudo nos comentários de PS3, a falta de diretrizes que orientem como pesquisadores devem proceder em relação a dados de pesquisa. Na visão de PS3, primeiramente, é necessário um entendimento da comunidade científica sobre o que são esses dados. Em um segundo momento, é preciso assegurar que dados continuem acessíveis ao longo do tempo, de modo a oportunizar o seu reuso em novas pesquisas. Sendo assim, cabe a reflexão sobre quais estratégias de preservação seriam úteis nesse contexto. Outro aspecto apontado por PS3 foi a adoção de medidas compulsórias que determinem a disponibilização de dados abertos gerados por pesquisadores em universidades. Nessa direção, PS3 aborda a questão da transparência da produção científica no ambiente acadêmico, conforme se pode observar no diálogo estabelecido entre PS3 e a entrevistadora:

Entrevistadora: *Você teria uma sugestão ou recomendação para sua universidade, caso a gestão de dados abertos seja institucionalizada?*

PS3: *Sim, primeiro fazer um ótimo treinamento com a gente dizendo sobre a importância de se entender o que são dados abertos, né? Essa questão que você colocou da preservação, pensar que daqui cem anos a gente pode ter alunos que estejam interessados nos dados que a gente está produzindo hoje, então, onde que a gente vai guardar esses dados? E segundo, eu acho que é indispensável que as universidades adotem algum modelo para que obriguem, de certa maneira, os professores a disponibilizarem as suas informações. São as informações coletadas em suas pesquisas em determinados repositórios para aumentar, inclusive, a transparência sobre o que é feito ali naquela universidade. Se não imediatamente, não sei, cinco, dez anos depois, alguma coisa no curto, médio, longo prazo, mas que de fato aumente o conhecimento e essa publicidade dos dados que são gerados ali naquelas universidades.*

Relacionado à falta de diretrizes institucionais está o aspecto da sobrecarga de trabalho em universidades vivenciada por pesquisadores, tais como PS10. De acordo com o seu entendimento, as universidades devem dispor de condições que viabilizem não somente a gestão de dados, mas em uma dimensão mais abrangente, ofereçam apoio à própria produção científica. PS10 mencionou que estabelece diversas parcerias em projetos de pesquisas internacionais, e que ao compará-los com o cenário brasileiro, esse último refere-se a uma produção que não dispõe de infraestrutura apropriada à pesquisa. Além disso, diversos coordenadores de projetos estariam sobrecarregados com tantos

procedimentos administrativos que devem ser efetuados institucionalmente. Por isso, PS10 propôs como alternativa a esse contexto, a disponibilidade de equipes especializadas para orientar e gerenciar parte do trabalho relacionada à gestão de dados abertos de pesquisa.

- Subcategoria “Diferenças disciplinares”

Diferenças disciplinares foram apontadas como aspecto influente na gestão de dados de pesquisa em universidades, tal como mencionado principalmente por PS3, PS6 e PS8. Sendo assim, PS3 chegou a cogitar como seriam as práticas em relação a dados produzidos em outras áreas, acreditando que no campo das Ciências naturais e da vida, a gestão de dados encontra-se em estágio mais avançado, no que se refere à adoção de estratégias de preservação.

Com relação às Ciências humanas, PS3 comentou que há muito o que se discutir nessa direção, uma vez que a submissão de projetos de pesquisas a comitês de ética ainda têm sido objeto de discordância entre os pares da área. Logo, falar sobre a disponibilização de dados de pesquisa parece ser algo bastante recente, porém, conforme disse PS3, é também uma questão a ser debatida urgentemente:

Eu acho que seria muito interessante ver se as práticas são diferentes a depender, por exemplo, dos colégios da Capes, né? A gente tem o colégio humanidades, colégio ciências da vida, colégio ciências naturais. Eu imagino que nesses colégios de ciência naturais e da vida, as práticas sejam mais avançadas do ponto de vista de preservação. Eu acho que seria muito interessante se a gente começasse esse debate pelo menos dentro das ciências humanas. Nas ciências humanas, eu acho que a gente está muitos passos atrás, porque a gente está discutindo a questão do comitê de ética, a gente não consegue se entender nem quanto ao comitê de ética, quiçá a disponibilidade de dados. Então, eu acho que esse é um debate urgente, especialmente quando a gente pensa do ponto de vista da preservação da memória, o que os nossos estudantes daqui há cem anos vão poder dizer sobre o que gente fazia hoje.

Ao falar sobre gestão de dados no ambiente acadêmico, PS6 explicou que as áreas de conhecimento vivenciam práticas de produção e disseminação científica diferenciadas, que possivelmente estariam associadas aos tipos de dados gerados. Por isso, sugeriu a realização de eventos que promovessem discussões a respeito desse tema em universidades. Logo, programas de pós-graduação disseminariam as informações discutidas a departamentos, de modo a considerar diferentes perspectivas de professores e estudantes.

De maneira semelhante, PS8 disse que a gestão de dados abertos na instituição em que atua necessita ser amplamente discutida, e que não sabe como essa temática se

apresenta em outras áreas. Com o propósito de exemplificar o aspecto abordado, PS8 disse que não conhece como dados são relacionados na produção de patentes, como devem ser armazenadas informações genéticas e nem como a legislação sobre dados pode ser aplicada nesse contexto.

Com base nas exposições sobre diferenças disciplinares no âmbito da gestão de dados, percebe-se a importância do tratamento e organização desses, antes de sua disponibilização pública, principalmente no que se refere ao cumprimento de requisitos éticos, tais como a anonimização de participantes da pesquisa. Para tanto, julga-se necessário que pesquisadores tenham acesso a recursos destinados à contratação de profissionais especializados em dados de pesquisa. Sobre esse aspecto, PS4 comentou:

Que fossem disponibilizados recursos específicos para a organização de dados de pesquisa, editais específicos para a contratação de pesquisadores que pudessem trabalhar para disponibilizar os nossos dados. Porque os dados não são uma questão de você pegar e colocar online, esses dados precisam ser trabalhados para serem disponibilizados. Na minha área, o pesquisador vai ter que ler esses diários de campo e verificar quais são os nomes, quais são os lugares que tem que ser trocados para que isso possa ser disponibilizado. É um material riquíssimo, mas vai demandar muito trabalho para ser disponibilizado. Não é uma questão de pegar a minha pasta e compartilhar. Esses dados precisam ser todos trabalhados para serem disponibilizados. Por razões éticas, eu tenho o compromisso com os meus interlocutores de não dizer quem eles são [...] E para isso, precisa ter recurso. Não adianta a universidade falar 'ah, põe no banco' e jogar para o pesquisador esse trabalho. Ela tem que falar 'oh, tem o banco aqui e a gente tem projetos para produzir esse banco'. Então, tem que contratar um grupo aí para fazer essa disponibilização dos dados.

4.2.3 Codificação seletiva e teoria substantiva relacionada à gestão de dados de pesquisa na área de Sociologia

Segundo Strauss e Corbin (2008), somente após as categorias principais forem integradas para a formação de um esquema teórico mais abrangente, é possível que resultados de pesquisas desenhem alguma forma de teoria. Sob tal perspectiva, a codificação seletiva pode ser entendida como o processo de integração e refinação de categorias. No que se refere à integração, e, especificamente, à escolha da categoria central por parte do pesquisador, os autores comentam:

A integração é um processo contínuo que ocorre com o tempo. Alguém poderia dizer que ela começa com a primeira parte da análise e não termina até a redação final. Como ocorre em todas as fases da análise, a integração é uma interação entre o analista e os dados. Dentro dessa interação, estão o formato analítico, que inclui não apenas quem o analisa, mas também a evolução do pensamento que ocorre com o tempo por meio da imersão nos dados e o conjunto cumulativo de resultados que foram registrados em memorandos e diagramas. Embora possamos encontrar nos dados dicas de como os conceitos se relacionam, elas só surgem depois que as relações são reconhecidas como tal pelo analista. Além disso, sempre que houver reconhecimento, há um certo grau de interpretação e de seletividade. [...] o primeiro passo na integração é decidir a categoria central. A categoria central (algumas vezes chamada de categoria básica) representa o tema principal da pesquisa. Embora a categoria central surja da pesquisa, ela também é uma abstração. Em um sentido exagerado, consiste de todos os produtos de análise, condensados em poucas palavras, que parecem explicar “sobre o que é a pesquisa” (STRAUSS; CORBIN, 2008, p. 144-145).

Devido às relações de centralidade estabelecidas com outras categorias, constatou-se, com base nas informações coletadas neste estudo, que a ética na pesquisa é a categoria principal para o entendimento das práticas de gestão de dados de pesquisa realizadas no contexto da Sociologia. Assim como é representado na Figura 30, pode-se destacar as seguintes relações observadas entre a categoria central e as demais categorias:

Figura 30 – Ética na pesquisa e sua relação com práticas de gestão de dados de pesquisa - Sociologia



Fonte: elaborado pela autora

- o tópico da Ciência aberta vem sendo discutido, principalmente, nos ambientes de convivência de pesquisadores da área de Sociologia, o que inclui grupos de pesquisa, departamentos acadêmicos e associações científicas. É nesse contexto que cientistas buscam entender como se procedem os aspectos éticos na pesquisa, em função das reais implicações acarretadas pelo fenômeno da Ciência aberta;
- considerando a predominância de instituições públicas no financiamento de pesquisas na área de Sociologia, tais como agências de fomento, alguns pesquisadores são solicitados a elaborar um Plano de gestão de dados, de modo a informar quais estratégias éticas serão adotadas em relação a conjuntos de dados. Essas dizem respeito, principalmente, à responsabilidade autoral e aos direitos de uso e compartilhamento de dados. Apesar de relevante, a exigência documental de um plano de gestão de dados de pesquisa ainda não é uma realidade amplamente difundida na área;
- a produção de dados na área de Sociologia revela-se por meio de uma diversidade tipológica. Frequentemente, pesquisadores trabalham com dados de temáticas sensíveis, os quais não poderiam ser expostos abertamente sem a observação de algum regramento ético. Diários de campo são um exemplo de material que pode trazer consigo muitos dados sensíveis, uma vez que pesquisadores costumam registrar a sua visão íntima sobre os fenômenos estudados;
- dados de temáticas sensíveis, que geralmente necessitam de maior atenção quanto a aspectos éticos, bem como dados de outra natureza, não costumam ser descritos por pesquisadores de acordo com padrões internacionais de metadados específicos

para a área de Sociologia. Logo, cada pesquisador adota o seu padrão para melhor descrever os dados que gera;

- a documentação sobre dados produzidos ao longo de pesquisas é relevante, principalmente, quanto ao registro de seu histórico, possibilitando assim, não somente a verificação da qualidade de dados, mas também a checagem de inconsistências metodológicas e éticas em pesquisas;
- o compartilhamento de dados no campo da Sociologia é visto, predominantemente, no contexto de grupos de pesquisa, ocorrendo por meio de canais informais. Nesse caso, é comum que o pesquisador estabeleça relações de confiança com quem deseja compartilhar os seus dados, avaliando, inclusive, a reputação e a conduta ética de outros cientistas. Ressalta-se, ainda, que o compartilhamento de dados é mediado pelo cumprimento de requisitos legais e éticos relacionados a conjuntos de dados, como a garantia de que a privacidade de participantes de pesquisas será resguardada;
- considerando que a preservação de dados abrange aspectos éticos relacionados à adoção de padrões de segurança de dados, principalmente no contexto daqueles que tratam de temáticas sensíveis, nota-se que pesquisadores na área de Sociologia adotam técnicas individuais para esse fim. Em geral, não há uma política institucional de preservação de dados ou não se exige de pesquisadores a elaboração de um plano de preservação, comprometendo o acesso contínuo a dados a longo prazo, bem como maximizando os riscos de perdas de dados;
- o reuso de dados dentre pesquisadores da área de Sociologia é predominante no contexto de redes de pesquisa, ocorrendo por meio de colaboração científica. Por conseguinte, são geradas muitas publicações em coautoria. Para que dados produzidos com o apoio financeiro de alguma agência de fomento possam ser reutilizados, geralmente, há acordos éticos nos quais se exige a citação da instituição fomentadora do estudo nas publicações derivadas;
- é fundamental que a gestão de dados abertos em universidades considere as diferenças disciplinares. A título de ilustração, convém mencionar que pesquisadores na área de Sociologia dispensam muita atenção em seus estudos à anonimização de dados e dos indivíduos participantes de pesquisas, principalmente, quando se

trata de temática sensível. Logo, essa é uma preocupação ética muito evidente para a área, carecendo de maiores debates sobre o tema.

À luz da literatura, a seção seguinte discutirá a categoria central – ética na pesquisa – a partir das relações estabelecidas com as demais categorias identificadas no estudo, que, por sua vez, referem-se a práticas de gestão de dados de pesquisa.

4.2.3.1 Ética na pesquisa e sua relação com práticas de gestão de dados de pesquisa

Nota-se que a Ciência aberta é um tópico de discussão presente no convívio de pesquisadores da área de Sociologia, abrangendo o contexto de grupos de pesquisa, departamentos acadêmicos e associações científicas. Relacionado a isso, está o aspecto da ética na pesquisa, sem o qual torna-se difícil aos cientistas compreenderem as reais implicações trazidas por esse fenômeno.

Sendo assim, entende-se a importância de a Ciência aberta ser amplamente discutida, não somente no meio científico, mas também no âmbito da sociedade como um todo. Tal como observado por Fecher e Friesike (2014), esse fenômeno está relacionado ao compartilhamento aberto de pesquisas e maior aproximação entre a comunidade científica e a sociedade. Trata-se de uma nova conjectura na qual diversos *stakeholders*, como cientistas, decisores políticos, editores científicos e analistas de sistemas, interagem em processos relativos à abertura de pesquisas e disseminação de conhecimento.

De outro vértice, Vicent-Saez e Martinez-Fuentes (2018) destacam que a Ciência aberta é apontada como um fenômeno que ocasionou mudanças socioculturais e tecnológicas, sendo essas fundamentadas no princípio da abertura do conhecimento científico. Para tanto, tem-se acentuado o desenvolvimento de plataformas abertas de acesso a dados, a criação de métodos de revisão por pares aberta e a realização de atividades que promovem maior interação entre pesquisadores e a sociedade.

Nesse contexto, Cribb e Hartomo (2010) argumentam em favor de uma produção científica mais aberta e democrática, ressaltando a necessidade de uma mudança cultural profunda no contexto científico. Para os autores, a democratização da ciência deve ser o pano de fundo para a compreensão da Ciência aberta. Desse

modo, a sociedade poderia contribuir para a produção científica, não somente portar-se como um receptor de informações.

Outro aspecto importante demonstrado no presente estudo foi que o financiamento de pesquisas na área de Sociologia é, predominantemente, oriundo de instituições públicas. Tendo como expoente as agências de fomento, é comum cientistas participarem de editais que oferecem recursos destinados à execução de pesquisas. Em linhas gerais, esse processo ocorre mediante a submissão de projetos os quais são avaliados e, se aprovados, passam a receber uma verba para esse fim.

À vista disso, alguns pesquisadores são solicitados por agências de fomento a elaborar um Plano de gestão de dados, no qual são descritas as principais ações em torno dos dados gerados na pesquisa. Nesse documento, aspectos éticos também são mencionados, principalmente no que se refere à indicação de responsabilidade e direitos de uso e compartilhamento sobre dados. Entretanto, apesar de sua importância, verificou-se que a exigência desse documento ainda não é uma realidade amplamente difundida no âmbito dos entrevistados, sendo até mesmo um termo desconhecido por muitos deles.

Ainda assim, preocupações éticas e alguns desafios nesse sentido são percebidos por pesquisadores, tal como foi apontado por Perrier e Barnes (2018), em estudo cujo objetivo foi identificar serviços desejáveis por pesquisadores para que a gestão de dados de pesquisa atenda a determinações de agências de fomento e editores de periódicos científicos. O estudo foi realizado a partir de quatro grupos focais com 28 participantes, incluindo professores da Universidade de Toronto, coordenadores de pesquisa, estudantes de pós-doutorado, bibliotecários e profissionais de tecnologia.

Perrier e Barnes (2018) concluíram que os entrevistados do estudo demonstraram preocupações com aspectos éticos da pesquisa, principalmente relacionados ao processamento de dados gerados a partir de perspectivas humanas. Sendo assim, tais dados deveriam ser cuidadosamente anonimizados e as informações confidenciais ser igualmente respeitadas, quanto aos termos de privacidade estabelecidos para a pesquisa.

O aspecto ético também foi uma das preocupações apontadas no estudo de Saeed e Ali (2019), cujo objetivo foi analisar práticas de gestão e compartilhamento de dados de pesquisa entre pesquisadores da *Aligarh Muslim University* (AMU). Para tanto,

o estudo aplicou questionários a 352 bolsistas de pesquisa no campo das Ciências da vida e das Ciências Sociais. Em relação a essas últimas, dentre os principais resultados obtidos na pesquisa, destacam-se dificuldades vivenciadas por pesquisadores em relação à manutenção da privacidade e a confidencialidade de dados produzidos.

No que se refere à coleta de dados, foi observada uma diversidade produtiva na área de Sociologia, ocasionada por diferentes técnicas aplicadas por pesquisadores. Trata-se de uma pluralidade na geração de dados, que inclusive, já foi percebida por autores como Sales e Sayão (2019), quando evidenciaram dois extremos na cadeia produtiva da ciência.

Segundo Sales e Sayão (2019), em um extremo encontra-se a produção e disseminação de dados homogêneos e padronizados, que ainda assim, necessitam de monitoramento e controle de processos executados na pesquisa. Em outro extremo, nota-se uma extensa cauda na ciência, caracterizada, sobretudo, pela diversidade de dados demonstrada por seus formatos, estrutura e domínio disciplinar. Em razão de tal heterogeneidade, os autores destacam a necessidade de gestão de dados de pesquisa, de modo a incluir especificidades socioculturais e tecnológicas trazidas por contextos disciplinares diversos.

Considerando a diversidade mencionada em relação aos tipos de dados de pesquisa, cabe destacar que a área de Sociologia produz, frequentemente, dados de temáticas sensíveis, os quais não poderiam ser disponibilizados sem a devida atenção a algum regramento ético.

De acordo com informações disponibilizadas no *site* do *Consortium of European Social Science Data Archives (CESSDA)*, muitos dados podem ser obtidos a partir de, ou se referem a participantes humanos. Se esses dados tratam sobre temáticas sensíveis, é importante assegurar que sua disponibilização à ampla audiência não ofereça riscos a direitos e liberdades de indivíduos que aceitaram participar de pesquisas. Dados sensíveis podem incluir informações relacionadas à origem étnica; escolha política; religião; orientação sexual; herança genética e estado de saúde.

Muitos dados sensíveis na área de Sociologia são registrados por pesquisadores em diários de campo, um instrumento pessoalíssimo no qual os mesmos expressam

a visão particular que possuem a respeito dos fenômenos estudados e dos eventos que os cercam. Ou seja, trata-se dos relatos íntimos de pesquisadores resultantes de observação participante e registrados em cadernos com certa periodicidade. Tais relatos podem incluir rotinas e descrições de indivíduos e ambientes, transcrição de entrevistas, memorandos de observação e até mesmo informações que expressam sentimentos vivenciados por pesquisadores em dado momento de um estudo (CARVALHO, 2018)

Semelhantemente, a relevante função exercida pelo diário de campo em pesquisas na área de Sociologia também foi identificada no estudo de Carvalho (2018) que objetivou identificar as práticas de produção/obtenção, compartilhamento e (re)uso de dados de pesquisa em Química, Antropologia e Educação. Para tanto, a autora realizou entrevistas semiestruturadas com docentes permanentes dos Programas de Pós-Graduação em Química (três pesquisadores), Antropologia (três pesquisadores) e Educação (quatro pesquisadores), da Universidade de Brasília (UnB).

Sendo uma referência para a coleta de dados antropológicos, um dos pesquisadores entrevistados por Carvalho (2018) explicou que o diário de campo registra não somente descrições de eventos e fenômenos, mas também as próprias percepções de pesquisadores e significados construídos no decorrer de suas pesquisas. Por se tratar, muitas vezes, de informações confidenciais e extremamente íntimas ao pesquisador, não é comum que tais instrumentos sejam compartilhados indiscriminadamente. Sobre tal aspecto o pesquisador entrevistado (Pesquisador 4) comentou:

Além de observações do que acontece, do que o pesquisador consegue entender no momento, o pesquisador faz sempre um diário de campo, pelo menos um, que é um referencial muito importante na coleta de dados do antropólogo. E nesse diário de campo, além de anotar o que é observado, o que é dito em descrições dos acontecimentos que o antropólogo quer entender, o antropólogo também reflete sobre o significado de tudo isso e sobre um tipo de comunicação que está conseguindo estabelecer com os sujeitos, que questões isso sugere e etc. Isso faz com que esse material do diário de campo seja um material que tem uma dimensão de intimidade que não é comum o pesquisador partilhar, com quem quer que seja. Eventualmente, quando os pesquisadores de um mesmo grupo de pesquisa, ou, às vezes, alunos com o professor, sob orientação do professor partilham parte desses dados, mas ninguém é obrigado a partilhar tudo. (PESQUISADOR 4, EM ENTREVISTA CONCEDIDA A CARVALHO, 2018)

Com base nas informações obtidas no estudo, dados de temáticas sensíveis, bem como os de qualquer outra natureza na área de Sociologia, não costumam ser descritos por pesquisadores conforme a adoção de padrões internacionais de metadados. Ou seja,

em geral, cada pesquisador utiliza o seu padrão para melhor organizar e recuperar os dados que produzem.

Porém, sabe-se que atualmente foram desenvolvidos alguns padrões internacionais de metadados destinados à área de Ciências sociais e Humanidades. É o caso do *Data Documentation Initiative*, associado à descrição de dados nas áreas das ciências sociais, comportamentais e econômicas. A versão DDI *Lifecycle* (ou DDI versão 3) foi proposta para lidar com conjuntos de dados complexos e abrange cada etapa de seu ciclo de vida. O *Statistical Data and Metadata Exchange* (SDMX) também se apresenta bastante útil no contexto de metadados estatísticos, principalmente, em razão da disponibilidade de orientações formais e diretrizes que oferece nesse campo específico (DIGITAL CURATION CENTRE – DCC⁶⁶)

Apesar de não ter sido criado especificamente para o campo das Ciências sociais e Humanidades, outro padrão de metadados adotado em diferentes áreas do conhecimento é o *Dublin Core*. Conforme observado nas entrevistas, nenhum pesquisador relatou lançar mão de quaisquer um desses padrões de metadados com a finalidade de descrição de dados que produz.

Sobre a importância da utilização de metadados na descrição de conjunto de dados, o ciclo de vida de dados criado pela GFBio⁶⁷ demonstra que metadados favorecem a localização de conjuntos de dados, possibilitando que esses sejam reutilizados por outros usuários. Além disso, metadados são fundamentais para elucidar o contexto de produção de dados, de maneira que explicam questões relacionadas à motivação para a criação de dados, à responsabilidade autoral de quem os gerou, ao local onde foram coletados e aos conteúdos que veiculam. Em sentido amplo, metadados contribuem para a descoberta de dados, tornando-os visíveis e acessíveis.

A documentação sobre processos e dados obtidos em pesquisas justifica-se, principalmente, devido ao registro do pesquisador do histórico de sua pesquisa, incluindo aspectos da produção, até a disponibilização e preservação de dados. Nota-se no campo da Sociologia, que esse processo documental também proporciona uma espécie de ateste

⁶⁶ Disponível em: <https://www.dcc.ac.uk/guidance/standards/metadata>

⁶⁷ Disponível em: <https://www.gfbio.org/pt/support/faqs>

da qualidade de dados, no qual pode-se verificar inconsistências metodológicas e éticas em pesquisas, e até mesmo, identificar improbidades nesse sentido.

De acordo com o projeto Orium⁶⁸ o fenômeno da Ciência aberta pode ampliar os níveis de qualidade e transparência de pesquisas, por meio da inclusão de práticas de acesso aberto ao longo do ciclo da pesquisa. Um dos caminhos apontados pelo projeto em direção à Ciência aberta é a chamada “pesquisa aberta”. Por sua vez, tal abordagem compreende que o compartilhamento dos processos de trabalho de pesquisa contribui para a avaliação aberta por terceiros.

A qualidade de dados disponibilizados abertamente também foi abordada no Guia de dados abertos, publicado em 2016, pela Agência para a Modernização Administrativa (AMA). No documento destacou-se que a abertura de dados não é alcançada apenas em sua disponibilização e divulgação por determinado período. Ao contrário, é preciso haver um compromisso institucional de disponibilização de dados a longo prazo, de modo a mantê-los atualizados, confiáveis e com qualidade de apresentação e acesso (fácil leitura, manuseio e uso).

Semelhantemente, no modelo do ciclo de vida de dados de pesquisa elaborado pela GFBio⁶⁹ pode-se mencionar a importância em se assegurar qualidade em relação a dados de pesquisa. Isso porque tal aspecto é um fator muito estimado para o reuso entre cientistas. Conseqüentemente, dados necessitam passar por monitoramento contínuo para que permaneçam acessíveis, legíveis e passíveis de interpretação.

Assim como foi demonstrado na pesquisa de Carvalho e Leite (2019) no que se refere ao compartilhamento de dados de pesquisa na área de Antropologia, notou-se no presente estudo que tal prática no campo da Sociologia ocorre, predominantemente, no âmbito de grupos de pesquisa, por meio de canais informais. Isso, baseado no estabelecimento de relações de confiança em que, geralmente, o cientista avalia a qualidade de dados e a reputação de outro pesquisador para proceder com o compartilhamento de dados.

No estudo de Carvalho e Leite (2019), observou-se que na área de Química, Antropologia e Educação, geralmente, o compartilhamento de dados mantinha-se

⁶⁸ Disponível em: <https://www.orion-openscience.eu/resources/open-science>

⁶⁹ Disponível em: <https://www.gfbio.org/pt/training/materials/data-lifecycle>

circunscrito ao contexto de grupos de pesquisa e às orientações entre professores e estudantes. Pesquisadores da Antropologia destacaram que tal prática deveria ocorrer em observância a critérios éticos preestabelecidos.

À luz dos aspectos mencionados, percebe-se a importância de canais informais na comunicação científica. Conforme explicou Mueller (2006), as publicações científicas disponíveis em canais formais, tais como os artigos de periódicos científicos, são tradicionalmente reconhecidas por pesquisadores, e o conhecimento nelas descrito é chancelado pelos pares. Porém, sabe-se que tais canais não são os únicos a possibilitarem a transmissão de conteúdos. No presente contexto, destaca-se que, frequentemente, dados de pesquisa são intercambiados por meio de canais informais.

De outro vértice, o aspecto ético relacionado ao compartilhamento de dados foi mencionado em documento elaborado pelo *Research Councils UK* (RCUK) (2016, p. 3). Nele, dados abertos de pesquisa são entendidos como aqueles que podem ser acessados, utilizados, modificados e disponibilizados de forma livre. Entretanto, algo é fundamental nesse contexto: o cumprimento de requisitos legais e éticos associados a conjuntos de dados. Sendo assim, é preciso assegurar a privacidade de participantes de pesquisas, considerar limitações previstas em termos de consentimento e preservar a integridade e confidencialidade de dados.

De modo semelhante, Gilmore *et al.* (2016) investigaram o tópico da abertura de dados à luz do compartilhamento de vídeos. Para tanto, os autores analisaram a biblioteca de dados digitais *Databrary* e expuseram alguns desafios éticos, técnicos e culturais vivenciados pela referida biblioteca digital, relacionados ao compartilhamento de vídeos de pesquisa.

Gilmore *et al.* (2016) ressaltaram que a criação de bibliotecas como a *Databrary* proporciona melhorias nos processos da ciência e avanços no campo da aprendizagem mediada por tecnologia. O projeto *Databrary* foi criado com o objetivo de promover o compartilhamento seguro de vídeos que tratam sobre ensino e aprendizagem e dispõe de mais de 500 membros pesquisadores e associados atuante em mais de 200 instituições.

Dentre os resultados obtidos no estudo de Gilmore *et al.* (2016), destaca-se que aspectos éticos são fundamentais no contexto de disponibilização de dados. Isso

porque o compartilhamento de vídeos demanda a adoção de políticas diferenciadas que garantam a privacidade de indivíduos expostos em pesquisas e que preservem a integridade de conteúdos que, posteriormente, serão reutilizados.

No que se refere a estratégias de preservação de dados de pesquisa, nota-se que cientistas da área de Sociologia adotam técnicas particulares para esse fim, não se tratando de padrões institucionalizados. Tais técnicas caracterizam-se, sobretudo, pelo armazenamento e preservação local e variam conforme os tipos de dados que são gerados na pesquisa. Em razão de ser uma prática individual e, geralmente, por não assegurar que dados sejam preservados a longo prazo, é visível o risco de perdas de dados, seja devido a desgastes físicos e temporais, seja pela descontinuidade de leitura, ocasionada pela obsolescência de mídias. Cabe destacar, ainda, que nenhum dos entrevistados no estudo afirmou ter elaborado algum Plano de preservação de dados.

No guia publicado em 2013 por grupo de trabalho da *DataOne*, intitulado *Data management guide for public participation in scientific research*, é apresentado o ciclo de vida de dados na trajetória de um projeto de pesquisa científica. Dentre as fases que compõe esse ciclo, está a preservação de dados, que, por sua vez, deve contemplar aspectos éticos como o estabelecimento de padrões de segurança de dados e a identificação de dados sensíveis. A preservação de dados abrange a elaboração de um plano de preservação de dados a curto e longo prazo, a realização de *backup* de dados, o armazenamento de dados em suportes estáveis e o depósito de dados em repositórios que prezem pela adoção de padrões de segurança de dados, principalmente, daqueles relacionados a temáticas sensíveis.

De outro vértice, no ciclo de vida de dados proposto pela GFBio⁷⁰, a preservação de dados é compreendida como o conjunto de procedimentos adotados para garantir a integridade de dados a longo prazo, de modo a mantê-los acessíveis, (recuperáveis e pronto para serem utilizados), íntegros (não suscetíveis a manipulações indevidas) e reutilizáveis. Especificamente no que se refere à preservação de dados no contexto digital, verificam-se técnicas relacionadas à migração e emulação de dados, além de arqueologia digital.

⁷⁰ Disponível em: <https://www.gfbio.org/pt/training/materials/data-lifecycle>

Sobre a preservação digital de dados, Martínez-Uribe e Fernández (2015) destacaram a importância de curadores demonstrarem competências no tratamento de dados dispostos em diferentes formatos, na descrição de metadados e na preservação digital, em especial, no contexto de repositórios científicos e na construção de ontologias e taxonomias relacionadas a conjunto de dados.

Neste estudo, também pode-se observar que o reuso de dados de pesquisa dentre os entrevistados da área de Sociologia concentra-se no âmbito de redes de pesquisa, dando-se por meio de colaboração científica. Como resultado, nota-se publicações científicas em coautoria.

De acordo com a *York University*⁷¹, o intercâmbio de dados abertos de pesquisa evita a duplicação de estudos, contribui para o aumento da reprodutibilidade de pesquisas e intensifica as oportunidades de colaboração científica. Para isso, é fundamental que tais dados sejam interoperáveis para viabilizar trabalhos colaborativos entre diversos pesquisadores e instituições.

Para Sayão e Sales (2015) o reuso de dados é bastante proveitoso para a continuação ou desenvolvimento de novos estudos. Isso porque dados podem ter tempo de vida maior que o período dos projetos de pesquisa que os geraram, podendo ser reutilizados.

Sob outra perspectiva, Silva (2019) afirmou que dados abertos de pesquisa podem ser definidos como aqueles que podem ser utilizados e compartilhados por quaisquer indivíduos, sem limitações de natureza técnica, ou legal. Porém, mesmo no contexto de abertura mencionado, alguns aspectos éticos devem ser considerados, como a devida atribuição de autoria de dados e a preferência de pesquisadores quanto à licença de direitos de reuso de dados. Essa, por sua vez, define legalmente o que é permitido e proibido em relação aos mesmos.

No presente trabalho, pesquisadores na área de Sociologia mencionaram que uma vez que recebem financiamento de alguma agência de fomento para a execução de uma pesquisa, devem cumprir com acordos éticos nos quais se exige a citação da agência em publicações resultantes do estudo. Tal exigência não diz respeito

⁷¹ Disponível em: <https://www.library.yorku.ca/web/open/overview/open-data/>

somente à referência de artigos a serem publicados, mas também a dados coletados no processo de produção científica.

Acrescente-se, por oportuno, que segundo o ciclo de vida de dados proposto por Rüegg *et al.* (2014) na descrição de um projeto de pesquisa tradicional, é fundamental que haja procedimentos de controle de qualidade sobre dados coletados, possibilitando que esses sejam posteriormente analisados. Todavia, os autores observam que, geralmente, o ciclo referido é finalizado após o pesquisador armazenar os dados que produziu em seus equipamentos pessoais, com pouca ou inadequada documentação sobre os dados.

Por conseguinte, dados são perdidos, na medida em que práticas necessárias a uma gestão efetiva tornam-se insuficientes perante o “dilúvio de dados” produzidos em diferentes áreas do conhecimento. Ademais, muitos pesquisadores não dispõem de capacitações para a gestão de dados nas instituições nas quais atuam. Logo, depreende-se que há uma reflexão tardia e não financiada no que se refere à importância do tema, demonstrando a necessidade da realização de debates como prestação de esclarecimentos sobre a gestão de dados de pesquisa.

De acordo com Wilkinson *et al.* (2016), uma “boa gestão de dados de pesquisa” contribui para a reutilização dos mesmos e integração de estudos científicos. Não há um consenso que afirme o que seja uma “boa gestão nesse sentido”, porém, os autores destacam que muitos dados não são explorados em sua totalidade, e por isso, é necessária a adoção de práticas que vão além da coleta e armazenamento apropriados. Nessa direção, a curadoria de dados a longo prazo possibilita a descoberta desses, viabilizando o seu reuso por diversos pesquisadores.

Um aspecto relevante para a compreensão da gestão de dados de pesquisa é a existência de diferenças disciplinares. Segundo Knorr-Cetina (1999), tais diferenças podem ser chamadas de “culturas epistêmicas” e se referem a divisões culturais na Ciência. Essas culturas refletem a diversidade de domínios disciplinares onde o conhecimento é gerado e são caracterizadas por práticas e padrões adotados por especialistas.

Tomando como pano de fundo as diferenças disciplinares apontadas por Knorr-Cetina (1999), percebe-se que alguns aspectos são fundamentais para a realização

de pesquisas em determinadas áreas. É o caso da atenção dispensada por pesquisadores de Sociologia em relação à anonimização de dados e de indivíduos nas pesquisas que realizam.

A anonimização de participantes de pesquisas fundamenta-se na preocupação ética de que pesquisadores adotam procedimentos para inviabilizar a identificação de indivíduos que consentiram participar de uma pesquisa científica. Trata-se de procedimentos que incluem a adoção de nomes fictícios para pessoas, instituições e até lugares.

Tais medidas demonstram-se extremamente importantes no contexto de pesquisas que versam sobre temáticas sensíveis à sociedade, como aquelas relacionadas a tipologias criminais e execução penal. Geralmente, dados dessa natureza demandam de pesquisadores o cumprimento de termos nos quais são estabelecidos algum nível de sigilo, ou anonimização de informações antes que o estudo seja publicado.

A título de ilustração sobre a importância do processamento técnico relacionado à anonimização de dados, convém destacar o estudo de Bellgard (2020) que elaborou um modelo conceitual de gestão de dados de pesquisa. Dentre os resultados obtidos, Bellgard (2020) destacou que dados coletados por meio de aparelhos hospitalares, como os de um tomógrafo, deveriam ser submetidos a um rigoroso trabalho de anonimização em diversas etapas de uma pesquisa. Além disso, é fundamental que sejam definidos os níveis de acesso a dados, a devida responsabilização autoral e os termos de utilização de dados em outros estudos.

4.2.4 Codificação aberta – Química

A presente seção apresentará a síntese dos principais aspectos abordados pelos pesquisadores de Química entrevistados no estudo. Descreverá as dimensões, categorias e propriedades identificadas por meio do processo da codificação aberta.

Semelhantemente aos procedimentos de anonimização adotados na pesquisa em relação aos pesquisadores de Sociologia entrevistados, todas as vezes em que pesquisadores em Química forem citados no texto, serão mencionados com o código “PQun^o”, em que “P” refere-se ao termo “pesquisador”, “Qu” ao termo “Química” e “n^o” ao número sequencial atribuído a cada pesquisador, conforme a ordem de realização das entrevistas.

4.2.4.1 Codificação aberta da entrevista com PQu1

Segundo PQu1, as discussões sobre Ciência aberta no campo da Química são recentes, assim como a preocupação sobre o acesso a dados de pesquisa. Quanto ao planejamento dos estudos que realiza, PQu1 disse que participa de uma série de projetos financiados por instituições de fomento à pesquisa. Porém, afirmou que ainda não teve experiência na elaboração de um Plano de gestão de dados de pesquisa, conforme explicitou:

Bom, normalmente, não! É uma coisa que realmente, talvez, deveríamos. Eu gostaria de ter avançado mais nesse aspecto.

No que se refere à coleta de dados, PQu1 apontou a relevância do caderno de laboratório físico ou digital para o registro dos experimentos que realiza rotineiramente. Além disso, os conjuntos de dados produzidos por PQu1 são descritos no âmbito de seu próprio laboratório, não sendo adotado nenhum tipo de padrão internacional de metadados. Quanto à documentação, PQu1 relatou que essa é considerada uma boa prática no contexto da produção científica, principalmente, por possibilitar a reprodução de experimentos.

Sobre o compartilhamento de dados de pesquisa, PQu1 destacou que a universidade em que atua tem um projeto de criação de repositório de dados de

pesquisa. Por meio desse, será possível atribuir metadados e relacionar o DOI com as publicações correspondentes, conforme apontou:

De fato, nós temos algum projeto que está sendo iniciado para ... digamos, a disponibilização de dados, não só através de repositórios, mas através de metadados em documentos. Mas é uma coisa que ainda está sendo iniciada [...] o nosso propósito é unirmo-nos a essa corrente de pensamento, dá um pouco de trabalho, mas não é muita coisa, você simplesmente coleta os dados todos, submete ao repositório e esse repositório é ligado com o DOI e a publicação correspondente.

PQu1 relatou que não teve experiência na elaboração de um Plano de preservação de dados de pesquisa. A preservação é realizada a partir de estratégias locais que incluem o armazenamento de dados em plataformas como o *google drive* ou o envio de dados no *e-mail*. Dessa forma, o risco de perda de dados e a obsolescência de mídias para a reprodução de dados é iminente, conforme apontado por PQu1:

Não temos um protocolo, digamos, de automatização para guardar os dados, embora como quase tudo é transferido via google drive, via email, os dados ficam sempre guardados [...] eu já perdi, às vezes, coleções de dados importantes por problemas com os backups. Naquela época era complicado, hoje em dia, é muito mais eficiente, né? Com o surgimento de todas essas tecnologias e se você tiver um pouco de cuidado, geralmente, pode garantir que quase nada é perdido.

Quanto ao reuso de dados na área em que atua, PQu1 relatou que é bastante comum pesquisadores citarem trabalhos publicados, não houve menção direta a dados, mas sim à informação publicada e devidamente referenciada. Finalmente, ao comentar sobre a gestão de dados de pesquisa, PQu1 abordou o aspecto das diferenças disciplinares, uma vez que é fundamental considerar a diversidade de dados e o manuseio desses no contexto disciplinar em que foram produzidos.

O Quadro 36 apresenta as dimensões, categorias e propriedades identificadas na codificação aberta, a partir da entrevista de PQu1.

Quadro 36 – Codificação aberta da entrevista com PQu1: práticas de gestão de dados de pesquisa

Dimensão	Categoria	Propriedade
Discussão sobre Ciência aberta	Discussões recentes	Tema atual
Planejamento	Financiamento à pesquisa	Editais
Coleta de dados	Caderno de laboratório	Registro de experimentos
Descrição de dados	Padrão local	Descrição estabelecida pelo pesquisador
Documentação de dados	Boas práticas de pesquisa	Registros no laboratório
Compartilhamento de dados	Criação de repositórios	Projeto de repositório de dados
Preservação de dados	Estratégias locais	Realizada pelos próprios responsáveis da pesquisa
Reuso de dados	Citação de trabalhos publicados	Referências
Gestão de dados abertos em universidades	Diferenças disciplinares	Dados de diferentes naturezas

Fonte: elaborado pela autora

4.2.4.2 Codificação aberta da entrevista com PQu2

Segundo PQu2, o tópico da Ciência aberta ainda é pouco discutido, apesar de ter conhecimento da realização de alguns fóruns de discussão na área, principalmente sobre acesso aberto. No que se refere ao planejamento das pesquisas que realiza, PQu2 afirmou que participa de editais para receber apoio financeiro de agências de fomento. Entretanto, relatou que é cada vez mais frequente a redução de verbas disponibilizadas por tais instituições, sendo fundamental realizar uma estimativa de gastos compatível com o orçamento disponível. Ademais, PQu2 disse que ainda não teve experiência na elaboração de um Plano de gestão de dados.

Quanto à coleta de dados, PQu2 relatou que esses resultam de colaboração científica e trabalho compartilhado dentro e fora da instituição em que atua, como se pode ver em sua fala:

E a maioria dos nossos projetos, posso dizer que todos, eles são interdisciplinares, então, em todos os meus projetos de pesquisa, eu conduzo em colaboração com grupos dentro da minha instituição, fora da minha instituição, no Brasil e fora da minha instituição, no exterior [...] a parte da própria caracterização, nós não temos aqui na nossa universidade equipamentos - todos os equipamentos necessários para fazermos toda a caracterização estrutural necessária da parte química. Então, a gente, invariavelmente, tem que mandar algumas amostras para alguns parceiros também para fazer isso, então é essa a sistemática que a gente utiliza.

De acordo com PQu2, conjuntos de dados obtidos são descritos conforme padrão local, principalmente, para atender à formatação exigida por revistas científicas

para a publicação de conteúdos. Para PQu2, documentar processos referentes a dados de pesquisa é fundamental para promover a divulgação científica e intensificar a interação entre universidade e sociedade. Nessa direção, PQu2 mencionou que as pesquisas documentadas no contexto universitário resultam em novos produtos e conhecimentos, que para ser transmitidos, devem se aproximar à linguagem utilizada pela sociedade.

Eu sou funcionário público, a gente faz pesquisa numa instituição pública, nós fazemos! [...] então, na pior das hipóteses, se a gente não gera um produto novo no final de uma pesquisa, a gente está gerando conhecimento, a gente está formando estudantes e a gente está formando novos profissionais, a gente está ampliando a participação da universidade no segmento social. Então, eu acho fundamental que nós consigamos divulgar o que a gente está fazendo, claro que, na maioria das vezes, tem que se dizer uma linguagem diferente da divulgação científica nos meios mais duros, nos meios científicos. Quando você vai conversar sobre o que está sendo feito dentro da universidade com a sociedade, você tem que se aproximar de uma linguagem que a sociedade vai compreender.

Ao discorrer sobre o compartilhamento de dados de pesquisa, PQu2 mencionou que o acesso aberto é prejudicado em razão da exigência de pagamentos para publicação de conteúdos científicos e para o acesso a esses. Comentou, ainda, que mesmo ao realizar uma palestra transmitida por uma plataforma *online*, com ampla audiência, não é possível assegurar acesso gratuito e irrestrito aos conteúdos.

Quanto à preservação de dados, PQu2 ressaltou que todos os membros da equipe de pesquisa a qual integra assina um termo de sigilo e confidencialidade a cada projeto de pesquisa desenvolvido. Tal medida é adotada com a finalidade de preservar os dados, além de prevenir a quebra do ineditismo por uma divulgação equivocada. Para PQu2, o reuso de dados de outro pesquisador seria algo eticamente reprovável. O que ocorre, muitas vezes, é a citação de trabalhos publicados os quais podem ser referenciados na literatura científica.

Por fim, ao falar sobre a gestão de dados de pesquisa no contexto de universidades, PQu2 alertou que no âmbito das ciências duras há dados que necessitam ser protegidos, não podendo ser amplamente divulgados por motivos econômicos. Nessa direção, PQu2 enfatizou a necessidade de sigilo a dados que podem resultar em inovação. Isso, principalmente em razão do interesse de indústrias em explorar economicamente produtos gerados em laboratórios de pesquisa. Logo, é fundamental que seja assegurada a proteção intelectual sobre dados de pesquisa.

O Quadro 37 apresenta as dimensões, categoria e propriedades identificadas na codificação aberta, conforme o relato de PQu2.

Quadro 37 – Codificação aberta da entrevista com PQu2: práticas de gestão de dados de pesquisa

Dimensão	Categoria	Propriedade
Discussão sobre Ciência aberta	Pouca discussão	Fóruns de discussão
Planejamento	Financiamento à pesquisa	Editais
	Recursos	Estimativa de recursos
Coleta de dados	Colaboração científica	Trabalho compartilhado
Descrição de dados	Padrão local	Descrição estabelecida pelo pesquisador
Documentação de dados	Divulgação científica	Interação entre universidade e sociedade
Compartilhamento de dados	Restrição de acesso	Acesso condicionado
Preservação de dados	Termo de sigilo e confidencialidade	Trabalho sob sigilo
Reuso de dados	Citação de trabalhos publicados	Referências
Gestão de dados abertos em universidades	Interesses econômicos	Restrições de divulgação de dados
	Sigilo de dados	

Fonte: elaborado pela autora

4.2.4.3 Codificação aberta da entrevista com PQu3

De acordo com PQu3, o tópico da Ciência aberta tem sido pouco discutido na área em que atua, sendo um assunto circunscrito ao departamento de sua universidade. Quanto ao planejamento das pesquisas que realiza, PQu2 disse que segue determinadas linhas de pesquisa, e a partir delas, elabora projetos e dimensiona os recursos necessários para atingi-los. Porém, relatou ainda não ter tido experiência na criação de um Plano de gestão de dados de pesquisa.

Ao falar sobre coleta de dados, PQu3 relatou que os obtêm, principalmente, a partir de levantamento em bases de dados internacionais. Os conjuntos de dados recebem uma descrição local do próprio pesquisador, não sendo adotados metadados específicos para essa finalidade. Além disso, PQu3 considerou que a documentação relacionada a dados de pesquisa é fundamental para se obter o histórico dos estudos realizados e, conseqüentemente, se ter acesso ao registro do conhecimento produzido

no laboratório. Ilustrando esse aspecto, PQu3 comentou a respeito do banco de dados sobre plantas que construiu ao longo dos anos:

A gente tem um banco de dados nosso, por exemplo, eu sei mais ou menos quantas plantas eu já investiguei nos últimos 20 anos, quantas plantas já foram investigadas com objetivos de composição química e atividade biológica. Então, isso eu sei porque a gente cataloga isso e, conseqüentemente, a gente tem um banco de dados das nossas publicações e também dos trabalhos que não foram publicados, foram só comunicados em congressos.

PQu3 relatou que já compartilhou dados de pesquisa após solicitação de revisores de uma revista científica. Segundo PQu3, é cada vez mais frequente que editores solicitem a pesquisadores que disponibilizem seus dados como material suplementar à publicação de artigos, possibilitando assim, a análise da veracidade de informações e dados submetidos a revistas.

PQu3 disse que ainda não teve experiência na elaboração de um Plano de preservação de dados, porém, destacou a importante função do caderno de laboratório para preservá-los. Os experimentos realizados são registrados por meio desse recurso impresso, que por sua vez, não sai do laboratório. Trata-se de uma estratégia em nível individual, e não institucional. PQu3 fez os seguintes comentários sobre o caderno de laboratório:

Ele [estudante que integra o grupo de pesquisa] chega lá às 7 horas da manhã no laboratório. E ele coloca o que vai fazer naquele dia, entendeu? Esse caderno aqui, se ele quiser levar, ele tira uma cópia, mas isso daqui fica no laboratório [...] aí você coleta planta, aí você prepara o material para encaminhar para um botânico e aí o botânico faz a identificação e confirma que aquela planta é aquela planta mesmo. Aí confirmou, aí a gente pega mais material e vai fazer a extração do que a gente quer, aí vai para o laboratório, aí depois faz análise química, tudo isso é anotado. Por que tudo isso é anotado? Porque quando a gente vai colocar isso no relatório para a comprovação que o aluno trabalhou aquilo ali, ou se a gente vai querer que seja publicado parcialmente esses resultados, ele vai ter que saber onde estão as informações e isso não dá para ficar na cabeça da gente, a gente termina se perdendo.

No que se refere ao reuso de dados, PQu3 afirmou que essa prática possibilita a realização de análises comparativas, de modo a justificar similaridades e diferenças entre dados. Além disso, ao comentar sobre a gestão de dados abertos em universidades, PQu3 apontou a importância do diálogo entre pesquisadores e universidade, de modo que essa ouça a opinião desses especialistas sobre o desejo o não do compartilhamento de dados. Outro aspecto abordado por PQu3 foi a perda do ineditismo ocasionada pela publicação antecipada de dados que tratam de inovação tecnológica. Nesse sentido, PQu3 alertou que outros pesquisadores poderiam publicar estudos com os seus dados, antes mesmo dele, que é o autor.

O Quadro 38 apresenta as dimensões, categoria e propriedades identificadas na codificação aberta, a partir da entrevista com PQu3.

Quadro 38 – Codificação aberta da entrevista com PQu3: práticas de gestão de dados de pesquisa

Dimensão	Categoria	Propriedade
Discussão sobre Ciência aberta	Pouca discussão	Departamento
Planejamento	Projeto de pesquisa	Linhas de pesquisa
Coleta de dados	Bases de dados	Levantamento em bases de dados
Descrição de dados	Padrão local	Descrição estabelecida pelo pesquisador
Documentação de dados	Bancos de dados	Histórico de pesquisas
Compartilhamento de dados	Periódicos científicos	Publicação em material suplementar
Preservação de dados	Estratégias locais	Realizada pelos próprios responsáveis da pesquisa
Reuso de dados	Análises comparativas	Estudos de similaridades e diferenças
Gestão de dados abertos em universidades	Diálogo	Consenso entre universidade e pesquisador
	Perda de ineditismo	Publicação antecipada de dados

Fonte: elaborado pela autora

4.2.4.4 Codificação aberta da entrevista com PQu4

De acordo com PQu4, a discussão sobre Ciência aberta necessita ser melhor explorada, em razão de muitas vezes a publicação de conteúdos científicos estar condicionada ao pagamento do autor, bem como ser também necessário o pagamento para obtenção de acesso a esses. PQu4 relatou que algumas revistas internacionais chegam a exigir milhares de dólares para a publicação de artigos em acesso aberto, o que inviabiliza sua publicação, devido à falta de recursos para essa finalidade.

PQu4 disse que ainda não teve experiência na elaboração de um Plano de gestão de dados. O planejamento das pesquisas que PQu4 realiza se baseia no financiamento obtido por meio de editais elaborados por agências de fomento. Porém, PQu4 destacou a redução nos valores dispostos nos editais, o prejudica a consecução de pesquisas, uma vez que a manutenção de equipamentos e a aquisição de reagentes são muito onerosas para o laboratório.

PQu4 enfatizou a obtenção de dados por meio de coletas sistemáticas, que prezam pela exatidão de resultados. De acordo com PQu4, a área das Ciências Exatas caracteriza-se pelo rigor na obtenção de medidas, apesar de haver uma margem de erro considerável na pesquisa. Dessa maneira, quando se extrapola a margem de erro estipulada, PQu4 avalia novamente os dados obtidos, em nível estatístico, e faz comparações na literatura científica para analisar o porquê de determinado resultado.

A descrição de conjuntos de dados é realizada localmente por PQu4. No que se refere à documentação sobre dados de pesquisa, PQu4 comentou que esse procedimento é importante por viabilizar a reprodutibilidade de pesquisas, principalmente, quanto aos métodos adotados. Já quanto ao compartilhamento de dados, PQu4 relatou que publica trabalhos completos em revistas e repositórios institucionais, ou seja, não há a publicação direta de dados.

PQu4 disse que ainda não teve experiência na elaboração de um Plano de gestão de dados. A preservação é realizada a partir de estratégias locais. No que se refere ao reuso de dados, PQu4 disse que é comum na área citar trabalhos publicados por outros pesquisadores, mas outra vez, destaca-se que não se trata da publicação direta de dados.

Ao falar sobre a gestão de dados abertos em universidades, PQu4 sugeriu que nessas instituições houvesse um setor responsável para lidar com a propriedade intelectual produzida por pesquisadores, o qual gerenciaria toda a questão da disponibilidade de dados. Nesse sentido, PQu4 disse:

Então, a minha sugestão para a nossa universidade é que o setor de propriedade intelectual esteja atento a esse tipo de disponibilidade dos dados abertos, que eu acho que a universidade faz isso.

O Quadro 39 apresenta as dimensões, categoria e propriedades identificadas na codificação aberta, de acordo com a entrevista com PQu4.

Quadro 39 – Codificação aberta da entrevista com PQu4: práticas de gestão de dados de pesquisa

Dimensão	Categoria	Propriedade
Discussão sobre Ciência aberta	Restrição de acesso	Acesso mediante pagamento
	Financiamento à pesquisa	Editais
Planejamento	Falta de estímulo ao pesquisador	Falta de recursos
Coleta de dados	Coleta sistemática	Exatidão na coleta de dados
Descrição de dados	Padrão local	Descrição estabelecida pelo pesquisador
Documentação de dados	Reprodutibilidade de pesquisas	Replicação de métodos
Compartilhamento de dados	Trabalhos publicados	Publicação em repositórios institucionais
Preservação de dados	Estratégias locais	Realizada pelos próprios responsáveis da pesquisa
Reuso de dados	Citação de trabalhos publicados	Referências
Gestão de dados abertos em universidades	Propriedade intelectual	Setor responsável pela disponibilização de dados abertos de pesquisa

Fonte: elaborado pela autora.

4.2.4.5 Codificação aberta da entrevista com PQu5

De acordo com PQu5, apesar de ser um tópico em evolução na comunidade científica, evidenciado em um contexto pandêmico no qual foi essencial o acesso a informações abertas, a Ciência aberta tem sido pouco discutida. Ainda se trata de um assunto incipiente na universidade em que atua.

Sobre o planejamento das pesquisas que realiza, PQu5 comentou a respeito das dificuldades para a obtenção de recursos no Brasil e que esses, por sua vez, são alcançados via participação em editais ou por meio de alguma parceria com empresas privadas. PQu5 informou que não teve experiência na elaboração de um Plano de gestão de dados. Porém, disse que teve conhecimento de que a universidade em que atua estaria alinhando a criação de um repositório de dados abertos de pesquisa, onde seriam unificadas informações de todos os grupos de pesquisa da instituição.

PQu5 comentou que, cada vez mais, projetos de pesquisa têm sofrido redução de verbas destinada para a sua execução, principalmente em relação ao financiamento oferecido por agências de fomento. Por isso, é imprescindível que o cientista realize

uma projeção de gastos aplicada à pesquisa, e que a mesma abranja possíveis custos ocasionados por fatores externos. De maneira ilustrativa, PQu5 comentou que custos de alguns de seus projetos de pesquisa foram alterados em razão da pandemia, especialmente, em decorrência de variações cambiais na importação de reagentes químicos e outros produtos.

Ao falar sobre a coleta de dados, PQu5 disse que, geralmente, lança mão de experimentos, os quais são todos registrados em cadernos de laboratórios digitais ou físicos. Em linhas gerais, são estabelecidas hipóteses as quais serão confrontadas a partir de experimentos. Dessa forma, poderão ser confirmadas ou refutadas.

Outro aspecto mencionado por PQu5 a respeito da coleta de dados foi o processo de validação de dados. Nessa direção, PQu5 comentou que os dados passam por tratamento estatístico rigoroso. Para tanto, é necessário repetir processos inúmeras vezes e realizar observações para fazer uma inferência mais segura em relação àquilo que se almeja comprovar.

A descrição de conjuntos de dados obtidos é realizada localmente por PQu5. Quanto à documentação de processos relativos a dados de pesquisa, PQu5 vê tal prática como sendo fundamental para a sistematização do conhecimento produzido, sem o qual a pesquisa não possui nenhum valor científico.

PQu5 comentou que ainda não teve experiência no compartilhamento de dados por meio de repositórios. Como citado anteriormente, a universidade em que atua está em processo de desenvolvimento de uma plataforma para esse fim. Ainda assim, PQu5 utiliza mídias sociais para a divulgação científica e compartilhar dados em linguagem mais próxima à da sociedade. Nessa direção, o grupo de pesquisa o qual integra tem um canal no *you tube* e uma conta no *Instagram* com objetivo principal de dar publicidade ao conhecimento científico produzido no contexto acadêmico.

No que se refere à preservação de dados de pesquisa, PQu5 disse que ainda não teve experiência na elaboração de um plano de preservação nesse sentido. Tal preservação é realizada por meio de estratégias locais, destacando-se o uso do caderno de laboratório físico ou digital para esse fim. Dada a sua importância no registro da pesquisa científica, esses cadernos não saem do laboratório, como se pode notar na fala de PQu5:

A gente tem o caderno de laboratório físico ou online. Mesmo quando o aluno deixa o laboratório, o caderno fica. Então, por exemplo, a gente tem uma dinâmica lá que caderno de laboratório não sai do laboratório. E o arquivo de caderno de laboratório, ele é geralmente via dropbox. Então, orientador tem acesso, aluno tem acesso. Então, quando eu estou falando de caderno de laboratório em pasta de dropbox, seria caderno virtual, vamos colocar dessa maneira: registro virtual. Aí eu vou ter dado bruto e dado tratado, né? Refinado.

Quanto ao reuso de dados, PQu5 disse é comum em sua área a citação de trabalhos publicados na literatura, principalmente para confrontar ou validar dados. Entretanto, não se trata do reuso de dados brutos. Ademais, ao falar sobre a gestão de dados abertos em universidades, PQu5 destacou que apesar de muito relevante, esse assunto precisa ser amplamente discutido, até mesmo para abranger as necessidades de cada área do conhecimento. Nessa direção, PQu5 sugeriu que no campo tecnológico deve haver entraves quanto ao compartilhamento de dados brutos, além de restarem dúvidas em relação ao momento em que o dado será divulgado (se a pesquisa estiver em andamento ou finalizada) e quanto aos critérios e/ou níveis de acesso aos dados.

Sobre a abertura de dados no âmbito de universidades, PQu5 ainda comentou a respeito da necessidade de dar transparência a esse processo, uma vez que se trata de um assunto que envolve não somente as perspectivas de instituições acadêmicas, mas também a de agências de financiamento à pesquisa, da própria sociedade e das especificidades inerentes a cada área do conhecimento.

PQu5 também comentou sobre a importância de proteção a dados a partir da adoção de protocolos, com o propósito de evitar sua apropriação indevida e a publicação antecipada por outros pesquisadores. Nesse sentido, PQu5 explicou que a velocidade de produção na universidade não é a mesma do setor produtivo. Se dados estiverem disponíveis sem nenhum tipo de proteção, tal como uma solicitação de patente, qualquer empresa poderia explorá-los como sendo seus, inclusive, antecipando a publicação de resultados.

O Quadro 40 apresenta as dimensões, categorias e propriedades identificadas na codificação aberta, conforme a entrevista com PQu5.

Quadro 40 – Codificação aberta da entrevista com PQu5: práticas de gestão de dados de pesquisa

Dimensão	Categoria	Propriedade
Discussão sobre Ciência aberta	Pouca discussão	Tema incipiente
Planejamento	Financiamento à Pesquisa	Editais
	Projeção de gastos aplicada à pesquisa	Variações de custos diante de fatores externos
Coleta de dados	Caderno de laboratório	Registros de experimentos
	Validação de dados	Rigor quanto à coleta de dados
Descrição de dados	Padrão local	Descrição estabelecida pelo pesquisador
Documentação de dados	Sistematização do conhecimento produzido	Atribuição de valor científico
Compartilhamento de dados	Divulgação científica	Uso de mídias sociais para a divulgação científica
Preservação de dados	Estratégias locais	Realizada pelos próprios responsáveis da pesquisa
Reuso de dados	Citação de trabalho publicado	Referência
Gestão de dados abertos em universidades	Diferenças disciplinares	Especificidades em relação à gestão de dados de pesquisa
	Transparência	Clareza nos processos relativos a dados de pesquisa
	Adoção de protocolos	Instrumento que mitiga riscos de apropriação indevida de dados

Fonte: elaborado pela autora.

4.2.4.6 Codificação aberta da entrevista com PQu6

De acordo com PQu6, o tópico da Ciência aberta tem sido amplamente discutido na área em que atua, principalmente, porque a comunidade científica que estuda a biodiversidade brasileira é extensa e tem intercambiado muitas informações sobre a temática em questão.

No que se refere ao planejamento das pesquisas que realiza, PQu6 relatou que recebe financiamento de agências de fomento para custear a produção científica, que, por sua vez, tem o custo bastante elevado, principalmente, devido à manutenção de aparelhos utilizados para a captação de espectros de substâncias. Além disso, PQu6 afirmou ter tido experiência na elaboração de um Plano de gestão de dados de pesquisa, atendendo à solicitação de uma agência de fomento à pesquisa.

Como PQu6 tem interesse de estudos na área de plantas e fungos, com o objetivo de localizar substâncias ativas para a composição de modelos biológicos, a

coleta de dados ocorre de acordo com a localidade em que se encontra a flora estudada. Nesse sentido, o acesso a plantas é fundamental para determinar a circunscrição geográfica na qual a pesquisa ocorrerá. PQu6 ainda especificou a maneira como se dá a coleta de plantas até chegar a proposição de um fármaco, como se pode observar em sua fala:

No meu caso, eu trabalho com química de produtos naturais. Basicamente, a gente trabalha com plantas e fungos. Então, há muitos medicamentos hoje em dia na quimioterapia, terapias no geral, que vieram de plantas. Então, esse estudo de base tem que ser feito, o que a gente faz? A gente coleta plantas, a gente prepara extratos, a gente avalia atividades nos mais diversos modelos in vitro e, posteriormente, in vivo, e a partir de um fracionamento daquele extrato, que é uma mistura grande de substâncias, a gente vai fazendo um trabalho de isolamento e chega em uma substância, uma estrutura química que compõem um protótipo de fármaco, por exemplo. Daí, a virar fármaco, vai muitos anos, mas o nosso é o trabalho de base mesmo.

Os conjuntos de dados coletados por PQu6 são descritos conforme os padrões exigidos por bancos de dados. No que diz respeito à documentação de processos relacionados a dados de pesquisa, PQu6 afirmou que essa prática é muito importante, não só para o registro e descrição de substâncias, mas também para dar continuidade ao conhecimento que foi construído ao longo do tempo. Para tanto, é preciso que dados estejam acessíveis a outros pesquisadores, conforme disse:

É o que a gente faz, é ciência básica. A gente descreve essa substância, se isso não está pavimentado, não permite dar continuidade para outros estudos. Então, os nossos dados têm que estar totalmente disponíveis para outros pesquisadores de áreas correlatas que possam pegar de onde a gente parou, para ir mais adiante.

PQu6 afirmou que teve experiência no compartilhamento de dados em dois bancos de dados internacionais de sequenciamento genético e descrição de material molecular, denominados, respectivamente: *GenBank* e *Global Natural Products Social Molecular Networking* (GNPS). Após a disponibilização de dados, qualquer pesquisador pode isolar características e analisar diferenças e semelhanças de genes.

A preservação de dados é realizada por PQu6, por meio do armazenamento de dados em discos rígidos, de plataformas em nuvem e de computador pessoal, da universidade que atua ou de alunos. Ainda sobre esse tópico, PQu6 disse que após solicitação de uma agência de fomento, elaborou um plano de preservação de dados no qual informou como seria estabelecida a guarda dos dados.

No que se refere ao reuso de dados, PQu6 reiterou que o aspecto da continuidade de estudos só é possível devido à reutilização de dados. Nesse sentido,

exemplificou que após ter acesso a dados publicados na década de 1980, pôde corrigir dados na literatura e atualizá-los a partir do uso de novas tecnologias aplicadas à pesquisa.

Por fim, ao discorrer sobre a gestão de dados abertos de pesquisa no contexto de universidades, PQu6 enfatizou que as maiores editoras de revistas científicas da área de Química cobram de autores taxas para publicar em acesso aberto, assim como a liberação a conteúdos é realizada somente mediante pagamento. Tais editoras tornam-se detentoras da propriedade intelectual sobre os conteúdos que veiculam. Entretanto, pesquisadores não recebem verbas destinadas para a publicação nesses termos. Por isso, PQu6 propôs que universidades e agências de financiamento à pesquisa assumam os encargos financeiros de publicações em acesso aberto, para que esse custeio não caia sob a responsabilidade de pesquisadores. A respeito disso, PQu6 mencionou que, certa vez, deveria pagar dois mil novecentos e cinquenta dólares para publicar um artigo em acesso aberto em um periódico internacionalmente reconhecido pela comunidade científica na área da Química.

O Quadro 41 apresenta as dimensões, categorias e propriedades identificadas na codificação aberta, conforme o relato de PQu6.

Quadro 41 – Codificação aberta da entrevista com PQu6: práticas de gestão de dados de pesquisa

Dimensão	Categoria	Propriedade
Discussão sobre Ciência aberta	Área de atuação	Discussões no campo de pesquisa
	Financiamento à pesquisa	Custos
Planejamento	Plano de gestão de dados	Plano solicitado por agência de fomento
	Localização geográfica	Coleta de dados conforme a aspectos geográficos
Coleta de dados	Localização geográfica	Coleta de dados conforme a aspectos geográficos
Descrição de dados	Padrão de bancos de dados	Estabelecido por plataformas internacionais
Documentação de dados	Continuidade de pesquisas	Estudos continuados a partir da publicação de dados
Compartilhamento de dados	Bancos de dados	Dados sobre sequenciamento molecular
Preservação de dados	Estratégias locais	Realizada pelos próprios responsáveis da pesquisa
Reuso de dados	Continuidade de pesquisas	Estudos continuados a partir do reuso de dados
Gestão de dados abertos em universidades	Propriedade intelectual	Responsabilização de encargos

Fonte: elaborado pela autora.

4.2.4.7 Codificação aberta da entrevista com PQu7

Para PQu7, a Ciência aberta é um tópico de pesquisa pouquíssimo discutido na área das ciências exatas. Ainda há muitas divergências na comunidade científica que envolvem direitos autorais e o compartilhamento irrestrito de dados de pesquisa, o que sugere maior aprofundamento sobre o tema.

No que se refere ao planejamento das pesquisas que realiza, PQu7 disse que costuma recorrer a editais de instituições públicas e ao estabelecimento de parcerias com empresas privadas, com a finalidade de angariar recursos para desenvolvê-las. PQu7 ainda comentou que não teve experiência na elaboração de um Plano de gestão de dados.

Quanto à coleta de dados, PQu7 destacou que obtém dados por meio de experimentos realizados em seu laboratório. Ou seja, é na bancada que sua equipe de pesquisa sintetiza novos materiais. Os conjuntos de dados coletados são descritos por PQu7 a partir de planilhas em *excel* e organizados conforme a destinação onde serão publicados: se irão gerar artigos, patentes, teses, entre outros canais. Assim também, PQu7 comentou que a documentação de processos relacionados aos dados coletados é muito importante para a organização administrativa do laboratório. Com esse propósito, PQu7 mantém pastas em seu computador com a produção científica de cada um de seus orientandos, de modo a registrar, sistematicamente, o desenvolvimento das pesquisas que coordena.

Com relação ao compartilhamento de dados, PQu7 disse que essa prática deveria ocorrer somente após estiver esgotado o interesse do autor dos dados em utilizá-los. Além disso, PQu7 comentou que ainda não teve experiência na elaboração de um Plano de preservação de dados. Tal preservação é realizada por PQu7, nada é proposto em nível institucional, como se pode notar em sua fala:

Não, não tem nada de institucional! Vai de cada pesquisador, não tem um lugar que a gente vai guardar. Então, acaba sendo no drive pessoal da gente e a gente faz o backup no HD externo.

Nesse contexto, PQu7 destacou que manuseia dados que requerem muito espaço para serem armazenados. Trata-se de dados obtidos por meio de instrumento de cromatografia tridimensional que capta compostos em determinadas amostras. Em uma dessas, é possível identificar cerca de 600 compostos que posteriormente são

analisados por meio do programa *Excel*. Em razão de esses dados exigirem extensa memória para serem armazenados, PQu7 disse que paga a plataforma *Google drive* para armazená-los em segurança.

PQu7 comentou que, é comum na área, pesquisadores citarem trabalhos já publicados, e não os dados de pesquisa, propriamente ditos. A referência a esses trabalhos permite a comparação de diferentes resultados. Além disso, PQu7 falou sobre a gestão de dados abertos em universidades e enfatizou a proteção à propriedade intelectual. Isso porque o foco das patentes no Brasil está no produto, não em ideias, sendo essas possivelmente geradas por meio de acesso a dados. Logo, PQu7 afirmou não se sentir seguro quanto a esse aspecto, como explicou:

Eu acho que tem que tomar cuidado a parte de dados abertos quanto à proteção de alguns trabalhos na área de propriedade intelectual. Porque como aqui no Brasil, a gente não pode patentear ideias como nos Estados Unidos, no Brasil não, no Brasil, ou a gente tem que ter um produto, fazer todo um documento daquele produto para patentear, seja um produto, seja um processo, ou uma outra coisa que tu podes patentear.

O Quadro 42 apresenta as dimensões, categorias e propriedades identificadas na codificação aberta, conforme a entrevista com PQu7.

Quadro 42 – Codificação aberta da entrevista com PQu7: práticas de gestão de dados de pesquisa

Dimensão	Categoria	Propriedade
Discussão sobre Ciência aberta	Pouca discussão	Tema incipiente
Planejamento	Financiamento à pesquisa	Editais
Coleta de dados	Experimentos	Dados obtidos em bancadas de laboratório
Descrição de dados	Padrão local	Descrição estabelecida pelo pesquisador
Documentação de dados	Organização do laboratório	Organização administrativa de equipe de pesquisa
Compartilhamento de dados	Pesquisas finalizadas	Compartilhamento de dados após a conclusão da pesquisa
Preservação de dados	Estratégias locais	Realizada pelos próprios responsáveis da pesquisa
Reuso de dados	Citação de trabalhos publicados	Referência
Gestão de dados abertos em universidades	Propriedade intelectual	Proteção à propriedade intelectual de ideias e produtos

Fonte: elaborado pela autora

4.2.4.8 Codificação aberta da entrevista com PQu8

De acordo com PQu8, a Ciência aberta é um tópico que vem sendo discutido na área, sobretudo, a partir de duas perspectivas: 1) o financiamento público da pesquisa, na qual a sociedade é o principal agente fomentador de pesquisas; 2) divulgação científica, em que se torna necessário adequar a linguagem científica para outra mais habitual à sociedade.

Para PQu8, quem custeia a produção científica no ambiente acadêmico é o povo, por meio do pagamento de impostos. Por isso, é razoável que os resultados obtidos sejam disponibilizados publicamente, conforme expressou:

Quem financia a pesquisa que a gente faz na universidade é o próprio povo, e o financiamento de pesquisa vem do regulamento de impostos. Então, a gente está desenvolvendo pesquisas, sendo financiado pelo governo federal, e esses recursos vem dos impostos do povo. Então, na minha opinião sincera aqui, eu acredito que as pesquisas têm que ser públicas, tudo que a gente faz aqui na academia tem que ser tornado público.

A respeito da divulgação científica, PQu8 disse que a produção em ciência tem sido publicada principalmente em revistas especializadas, sendo revisada por pares para avaliação de mérito científico. Todavia, trata-se ainda de um conhecimento circunscrito à universidade que não é acessível a toda a população. Outro aspecto que para PQu8 caberia maiores discussões é a linguagem utilizada na publicação científica, que a seu ponto de vista, deveria ser simplificada e mais geral para tornar a comunicação com a sociedade mais eficaz. Nessa direção, PQu8 explicou:

A linguagem que a gente usa em uma publicação científica é um negócio muito técnico, só o meu colega de sala que vai entender o que eu estou falando. Quem trabalha na mesma linha de pesquisa vai entender o que eu estou falando, mas a população em geral, não. Então, já está havendo também essa preocupação em tornar os resultados mais acessíveis, em uma linguagem mais usual mesmo, que todo mundo possa entender [...] a gente não torna isso, não leva isso para a sociedade, o que a gente deveria fazer? Pegar os nossos resultados aqui e levar isso para a sociedade usando uma linguagem mais simplificada e mais geral. Então, essa parte a gente está falhando mesmo e isso a gente precisa fazer: melhorar essa comunicação com a sociedade.

PQu8 disse que ainda não teve experiência na elaboração de um Plano de gestão de dados de pesquisa e que não possui nenhum instrumento institucional que o auxilie nessa direção. O planejamento relacionado as suas pesquisas é realizado em parceria com os seus orientandos. Quanto à coleta de dados, PQu8 comentou que tem se dedicado à investigação de formas de cura para determinadas doenças. Para tanto, costuma coletar amostras de materiais biológicos de pacientes humanos,

presentes em urina, sangue, saliva, entre outros. Os participantes da pesquisa assinam um termo de consentimento em que se registra a doação da amostra biológica para o estudo.

PQu8 também relatou que a descrição de conjuntos de dados é realizada de modo individual, não havendo a adoção de um padrão de metadados uniforme. Todavia, destacou a necessidade de uniformização de metadados espectrométricos para que os mesmos possam ser utilizados por qualquer pesquisador no mundo. Além disso, comentou sobre a importância de se documentar todos os processos relativos à obtenção de dados, uma vez que tal medida é essencial para a comprovação de que o dado é fidedigno.

PQu8 ainda comentou sua experiência no compartilhamento de dados abertos relativos a uma pesquisa que utilizou amostras de sangue de paciente para o desenvolvimento de um método para o diagnóstico de doença encefálica. Desse modo, PQu8 disponibilizou publicamente dados originais e dados de equipamentos em uma plataforma criada pela universidade em que atua.

No que se refere à preservação de dados, PQu8 afirmou que ainda não teve experiência na elaboração de um Plano de preservação de dados. Tal preservação é feita por meio de estratégias locais no laboratório, não sendo realizada em nível institucional. Com relação ao reuso de dados, PQu8 disse que é comum pesquisadores reutilizarem dados das pesquisas que conduz, principalmente dados espectrométricos de moléculas. Nessa direção, tal prática viabiliza a proposição de novos estudos, inclusive, sob perspectivas diferentes da pesquisa que deu origem aos dados.

Finalmente, ao falar sobre a gestão de dados abertos em universidades, PQu8 mencionou que o pesquisador disposto a compartilhar seus dados poderia ser recompensado de alguma maneira. As recompensas poderiam incluir pontos na progressão do pesquisador, e, conseqüentemente, aumento salarial. Nesse sentido, PQu8 explicou:

O que vem na cabeça agora é uma valorização de quem coloca os dados públicos [...] se o pesquisador for valorizado para ele tornar esses dados públicos, aí eu acho que a gente vai lá e vai colocar os dados públicos. Se eu ganhar um pontinho aqui na minha progressão, por exemplo, a cada dois anos o professor ele progride na carreira e aumenta o salário, então poderia ser um exemplo aqui.

O Quadro 43 apresenta as dimensões, categorias e propriedades identificadas na codificação aberta, conforme a entrevista com PQu8.

Quadro 43 – Codificação aberta da entrevista com PQu8: práticas de gestão de dados de pesquisa

Dimensão	Categoria	Propriedade
Discussão sobre Ciência aberta	Financiamento público da pesquisa	A sociedade financia a pesquisa
	Divulgação científica	Adoção de linguagem utilizada pela sociedade
Planejamento	Falta de auxílio institucional	Planejamento realizado pelo próprio pesquisador
Coleta de dados	Amostras de materiais biológicos	Coleta autorizada por participantes da pesquisa
Descrição de dados	Padrão local	Descrição estabelecida pelo pesquisador
Documentação de dados	Comprovação de dados	Comprovação da fidedignidade de dados
Compartilhamento de dados	Plataforma de dados abertos	Compartilhamento de dados em plataforma criada pela universidade
Preservação de dados	Estratégias locais	Realizada pelos próprios responsáveis da pesquisa
Reuso de dados	Proposição de novos estudos	Utilização de dados em novos estudos
Gestão de dados abertos em universidades	Sistema de recompensas	Recompensas a pesquisadores que compartilham dados abertos

Fonte: elaborado pela autora.

4.2.4.9 Codificação aberta da entrevista com PQu9

Segundo PQu9, a Ciência aberta é um tópico de estudo discutido há mais de 20 anos e que tem assumido um viés comercial, em razão de muitas revistas científicas cobrarem pela publicação de conteúdos em acesso aberto. Sob essa ótica, o autor ou quem o patrocina custeia tal abertura, enquanto as revistas tornam-se proprietárias da informação a ser publicada.

O planejamento das pesquisas de PQu9 é realizado com base no financiamento obtido principalmente por agências de fomento, conforme estabelecido em editais. Com o recebimento de recursos, é possível aprimorar a estrutura do laboratório que coordena, assim como a qualidade dos resultados obtidos. Ademais, PQu9 comentou que ainda não teve experiência na elaboração de um Plano de gestão de dados de pesquisa.

Geralmente, os dados coletados por PQu9 são resultados de estudos longitudinais relacionados à ciência básica, durando cerca de uma década. Os conjuntos de dados obtidos são descritos localmente por PQu9. Não se adota um padrão de metadados, apesar da existência de normas internacionalmente aceitas que parametrizem como deve ocorrer a produção de dados.

Ao falar sobre a importância da documentação em relação a dados de pesquisa, PQu9 comentou que essa prática é fundamental para a verificação da fidedignidade de dados e torna-se um recurso extremamente útil ao pesquisador, quando o mesmo é solicitado a comprovar o trabalho realizado. Já quanto ao compartilhamento de dados, PQu9 disse que a universidade em que atua tem uma proposta de criação de repositório de dados, principalmente em razão de exigências de organizações de pesquisas, mas que ainda não é uma realidade.

PQu9 não teve experiência na elaboração de um Plano de preservação de dados de pesquisa, todavia, comentou sobre a importante função do caderno de laboratório para a preservação de dados. O uso desse recurso impresso, que cada vez mais tem migrado para o meio digital, armazena o que é produzido no laboratório. Cotidianamente, estudantes registram o que produzem e todo o material permanece naquele local. PQu9 destacou, ainda, que não há estrutura informatizada na instituição com a finalidade de preservação de dados.

Para PQu9, é comum na área, pesquisadores fazerem a citação de trabalhos publicados, e não de dados. Além disso, PQu9 falou sobre a gestão de dados abertos na universidade em que atua. Nesse sentido, sugeriu que fosse criado um setor responsável por esse processo na instituição, juntamente ao Centro de processamento de dados da universidade. Tal medida seria importante, na medida em que o volume de dados de pesquisa produzidos é extenso, além de requerer a destinação de recursos e de infraestrutura tecnológica para esse fim.

Eu penso que deveria ser criado um setor específico dentro do centro de processamento de dados para tratar dessa questão dos dados gerados pela pesquisa e para a criação dos repositórios de uma maneira atualizada e formal, como deve ser [...] esse é um dos problemas que a gente tem aqui, porque na [nome da universidade que atua] o volume de dados gerados em pesquisa é muito grande. Então, nós teríamos que ter um setor destinado para isso, inclusive, com recursos de servidores computacionais que pudessem dar conta de armazenar o dado para a instituição poder disponibilizar de acordo com o seu regramento.

O Quadro 44 apresenta as dimensões, categorias e propriedades identificadas na codificação aberta, conforme a entrevista com PQu9.

Quadro 44 – Codificação aberta da entrevista com PQu9: práticas de gestão de dados de pesquisa

Dimensão	Categoria	Propriedade
Discussão sobre Ciência aberta	Acesso pago	Acesso a publicações mediante pagamento
Planejamento	Financiamento à pesquisa	Editais
Coleta de dados	Estudos longitudinais	Dados coletados ao longo de décadas
Descrição de dados	Padrão Local	Descrição estabelecida pelo pesquisador
Documentação de dados	Comprovação de dados	Verificação da fidedignidade de dados
Compartilhamento de dados	Criação de repositórios	Proposta de criação de repositório de dados
Preservação de dados	Estratégias locais	Realizada pelos próprios responsáveis da pesquisa
Reuso de dados	Citação de trabalhos publicados	Referência
Gestão de dados abertos em universidades	Centro de processamento de dados	Setor responsável pela gestão de dados na universidade

Fonte: elaborado pela autora.

4.2.4.10 Codificação aberta da entrevista com PQu10

De acordo com PQu10, tem se discutido o tópico da Ciência aberta em nível nacional e internacional, sobretudo, com o propósito de se questionar o atual modelo de publicação científica. Nele, observa-se que grandes editores e jornais tornaram-se proprietários da produção científica, exigindo o pagamento de taxas para a publicação e acesso a conteúdos. Todavia, o fenômeno da Ciência aberta tem se fortalecido e argumentado em favor de que o conhecimento deve ser disponibilizado publicamente, sem restrições de pagamento. Nesse sentido, PQu10 comentou que pesquisas realizadas com recursos públicos deveriam também ser abertas.

O planejamento das pesquisas que PQu10 realiza são baseados na disponibilidade de recursos oferecidos por agências de fomento, conforme previsto em editais. Entretanto, PQu10 destacou que observa uma diminuição contínua de recursos destinados à pesquisa, o que tem impactado negativamente a produção

científica. Isso porque as linhas de financiamento oscilam muito no Brasil, dificultando não somente a execução de um projeto de pesquisa, mas também a manutenção da estrutura de pesquisa, que inclui os custos com laboratório, contratação e treinamento de equipes. Além disso, PQu10 comentou que ainda não teve experiência na elaboração de um Plano de gestão de dados, apesar de lidar rotineiramente com toda a gestão científica e técnico-administrativa do laboratório.

Quanto à coleta de dados, PQu10 destacou a importante função do caderno de laboratório para o registro de experimentos. Geralmente, esse recurso permanece no laboratório, saindo do ambiente apenas de modo eventual para o cumprimento de análises específicas. Os conjuntos de dados obtidos são descritos localmente por PQu10, em seu laboratório. Além disso, PQu10 acha imprescindível a documentação de processos relativos à produção de dados, principalmente para se obter uma pesquisa com qualidade, que inspire confiança. Definindo-se como um cientista experimentalista, fez a seguinte afirmação “se eu não tenho confiança no dado, não tem pesquisa, eu sou experimentalista!”

PQu10 disse que não teve experiência em disponibilizar os dados de pesquisa de forma amplamente aberta, até mesmo por não dispor de recursos financeiros destinados a esse fim. Geralmente, os dados de suas investigações são compartilhados entre os membros do grupo de pesquisa o qual integra. Nessa direção, PQu10 afirmou que adota em seu laboratório um critério de gestão cooperativo, e não competitivo.

PQu10 ainda não teve experiência na elaboração de um Plano de preservação de dados. Entretanto, algumas estratégias são realizadas no âmbito do laboratório, tais como a manutenção de equipamentos e computadores. Outro aspecto apontado por PQu10 foi o condicionamento térmico necessário para a preservação de resultados de pesquisas, e, conseqüentemente, de dados obtidos. Para tanto, é fundamental ter à disposição aparelhos de ar condicionado e equipamentos robustos capazes de operar em diferentes temperaturas.

Quanto ao reuso de dados, PQu10 mencionou a citação de trabalhos publicados na literatura científica, não especificamente de dados de pesquisa. Sobre isso, PQu10 explicou:

E quando se cria uma coisa nova, a gente precisa ter um padrão de comparação. Um dado de comparação de um dado para outro é um dado que está disponível. Como é que você sabe que uma

coisa é maior, menor ou pior? Comparando com o que foi feito. Então, o que eu tento é estar sempre atualizado na literatura. Não é fácil, porque hoje o volume de informação que o mundo está produzindo é muito grande.

Ao falar sobre a gestão de dados abertos no contexto de universidades, PQu10 destacou que não haveria condições de dados de pesquisas em andamento serem disponibilizados, por se tratar de dados não finalizados ou de resultados prematuros. Sendo assim, esse tipo de dado é passível de ser alterado e maturado, até a obtenção de um resultado concreto. Se publicado antes do previsto, pode não ter sido suficientemente averiguado quanto ao nível de qualidade que apresenta, ou até mesmo apresentar inconsistências que sugerem erros para a aplicação em outros estudos. A respeito disso, PQu10 explicou:

Qualquer estratégia de pesquisa, ela pressupõe que você, para concretizar a tua ideia, concretizar o teu produto, você precisa maturar esses seus dados. Então, você não pode ter um dado agora e todo mundo pega isso aí e faz como quer, porque senão você não consegue nem checar se aquele teu dado é correto. Por exemplo, é lógico que a gente já fez medida errada no laboratório, por isso que eu tenho um padrão de controle, então, eu não posso fazer uma descoberta agora e amanhã ela já estar para um acesso público [...] então, para a gente ter certeza, a gente precisa de um tempo para poder estar com aquilo, vamos dizer, robusto, concretizado.

O Quadro 45 apresenta as dimensões, categorias e propriedades identificadas na codificação aberta, conforme a entrevista com PQu10.

Quadro 45 – Codificação aberta da entrevista com PQu10: práticas de gestão de dados de pesquisa

Dimensão	Categoria	Propriedade
Discussão sobre Ciência aberta	Acesso público a dados	Dados obtidos por meio de financiamento público
	Financiamento à pesquisa	Editais
Planejamento	Diminuição contínua de recursos	Oscilação nas linhas de financiamento de pesquisas
	Infraestrutura da pesquisa	Falta de infraestrutura da pesquisa no Brasil
Coleta de dados	Caderno de laboratório	Registro de experimentos
Descrição de dados	Padrão local	Descrição estabelecida pelo pesquisador
Documentação de dados	Confiança nos dados	Qualidade dos dados
Compartilhamento de dados	Grupo de pesquisa	Entre membros do grupo de pesquisa
Preservação de dados	Estratégias locais	Realizada pelos próprios responsáveis da pesquisa
Reuso de dados	Citação de trabalhos publicados	Referências
Gestão de dados abertos em universidades	Resultados prematuros	Dados de pesquisa em andamento

Fonte: elaborado pela autora.

4.2.4.11 Síntese da codificação aberta

O processo de codificação aberta nas entrevistas resultou em uma primeira descrição das práticas de gestão de dados utilizadas por pesquisadores no campo da Química. Nesse sentido, foram identificadas 61 categorias abertas, dispostas em nove dimensões. Tais categorias passaram por processo de codificação axial, convertendo-se em outras nove categorias axiais, que, por sua vez, se desdobraram em dez subcategorias, as quais podem ser observadas no Quadro 46.

Quadro 46 – Síntese da codificação aberta - Química: práticas de gestão de dados de pesquisa

Dimensão	Categoria aberta	Categoria axial	Subcategoria axial
Discussão sobre Ciência aberta	Discussões recentes	Ciência aberta	Restrições de acesso
	Pouca discussão		
	Restrição de acesso		
	Área de atuação		
	Financiamento público da pesquisa		
	Acesso pago		
	Acesso público a dados		
Planejamento	Financiamento à pesquisa	Planejamento da pesquisa	Financiamento à pesquisa
	Recursos		
	Projeto de pesquisa		
	Falta de estímulo ao pesquisador		
	Projeção de gastos aplicada à pesquisa		
	Plano de gestão de dados		
	Falta de auxílio institucional		
	Diminuição contínua de recursos		
Infraestrutura da pesquisa			
Coleta de dados	Caderno de laboratório	Coleta de dados	Caderno de laboratório
	Colaboração científica		
	Bases de dados		
	Coleta sistemática		
	Validação de dados		
	Localização geográfica		
	Experimentos		
	Amostras de materiais biológicos		
Estudos longitudinais			
Descrição de dados	Padrão local	Descrição de dados	Padrão local
	Padrão de bancos de dados		

Documentação de dados	Boas práticas de pesquisa	Documentação de dados	Comprovação de dados
	Divulgação científica		
	Bancos de dados		
	Reprodutibilidade de pesquisas		
	Sistematização do conhecimento produzido		
	Continuidade de pesquisas		
	Organização do laboratório		
	Comprovação de dados		
Confiança nos dados			
Compartilhamento de dados	Criação de repositórios	Compartilhamento de dados	Repositórios/ Plataformas de dados abertos
	Restrição de acesso		
	Periódicos científicos		
	Trabalhos publicados		
	Divulgação científica		
	Bancos de dados		
	Pesquisas finalizadas		
	Plataforma de dados abertos		
Grupo de pesquisa			
Preservação de dados	Estratégias locais	Preservação de dados	Estratégias locais
	Termo de sigilo e confidencialidade		
Reuso de dados	Citação de trabalhos publicados	Reuso de dados	Trabalhos publicados
	Análises comparativas		
	Continuidade de pesquisas		
	Proposição de novos estudos		
Gestão de dados abertos em universidades	Diferenças disciplinares	Gestão de dados abertos em universidades	Diferenças disciplinares
	Interesses econômicos		
	Diálogo		
	Perda de ineditismo		
	Propriedade intelectual		Propriedade intelectual
	Transparência		
	Sistema de recompensas		
	Resultados prematuros		
Sigilo de dados			
Adoção de protocolos			

Fonte: elaborado pela autora

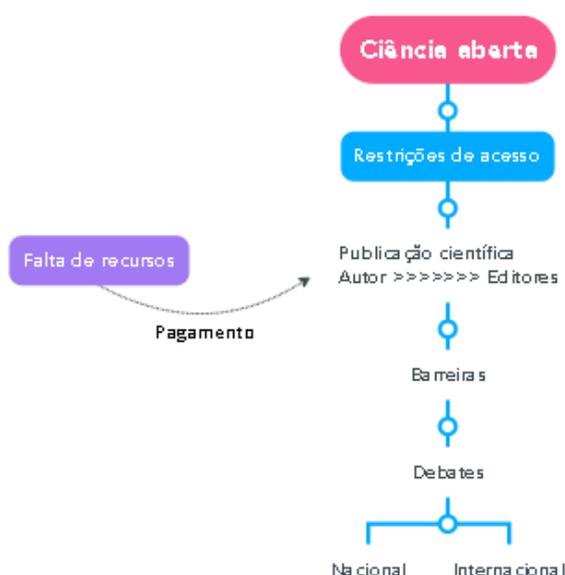
4.2.5 Codificação axial - Química

Esta seção apresentará relações identificadas entre as categorias e subcategorias obtidas no estudo, com base no processo de codificação axial desenvolvido.

4.2.5.1 Categoria “Ciência aberta”

Como representada na Figura 31, a categoria “Ciência aberta” foi desenvolvida em termos da subcategoria “Restrições de acesso”. Nesse sentido, entrevistados apontaram um viés relacionado ao acesso aberto de conteúdos científicos: o pagamento de taxas de publicação realizado por pesquisadores a editores. Todavia, tais entrevistados também afirmaram que não dispõem de recursos suficientes para custear esse tipo de publicação. Em linhas gerais, esse cenário resulta em barreiras ao amplo acesso ao conhecimento científico que é produzido, e demonstra a necessidade dessa temática ser discutida em âmbito nacional e internacional.

Figura 31 – Categoria Ciência aberta - Química



Fonte: elaborado pela autora

- Subcategoria “Restrições de acesso”

Com base nos relatos de entrevistados, em especial (PQu2, PQu3, PQu5 e PQu7), observou-se que a Ciência aberta ainda é um tópico pouco discutido na área da Química, mesmo com a realização de fóruns e outros eventos acadêmicos. Segundo PQu7, trata-se de um tema recente em que há vários pontos de divergência na própria comunidade científica relacionados, principalmente, com direitos autorais e o compartilhamento irrestrito de dados de pesquisa.

Ainda assim, nota-se que a Ciência aberta é um conceito em evolução na área de Química, tendo sido demonstrada a sua relevância principalmente no contexto pandêmico recente, no qual foi vital a maximização de informações, o compartilhamento de dados abertos e a minimização de restrições de acesso ao conhecimento produzido.

É nesse contexto que entrevistados como PQu4, PQu8, PQu9 e PQu10 destacaram restrições de acesso a dados, mesmo ao publicarem conteúdos científicos por vias abertas. Em muitas situações, esse tipo de publicação ocorre mediante o pagamento do autor, o que consequentemente restringe não só a produção científica, mas também o acesso à mesma. Nesse sentido, alguns títulos de periódicos internacionais chegam a cobrar milhares de dólares para a publicação de artigos em acesso aberto, o que geralmente se torna inviável para o autor brasileiro, devido à falta de recursos destinados para isso.

Desse modo, PQu4 comentou sobre taxas para a publicação de artigos em acesso aberto, lembrou a importância da base de dados de periódicos da Capes como um canal de acesso gratuito a conteúdos científicos e ainda destacou que autores cedem direitos autorais a revistas, que por seu turno, passam a ser detentoras de informações e dados nelas veiculados:

Olha, a gente tem algumas revistas que trabalham com publicações abertas, mas é tudo pago. No caso, revistas internacionais, você opta por publicar seu trabalho de forma aberta, mas ele te cobra, tá? Os nossos trabalhos, eles são publicados nessas revistas, mas a gente tem o acesso a partir da Capes, né? Periódicos da Capes. Então, o próprio Ministério da educação paga para a gente ter o acesso. É como se fosse as revistas que nós temos, se você paga, você tem acesso, se você não paga, você não tem acesso. Então, às vezes, se eu não estiver conectado no meu computador, eu não tenho acesso aos meus próprios trabalhos, entendeu? Eu tenho que está conectado à plataforma Capes e aí eu tenho acesso diretamente. E aí a gente tem um jornal e a gente paga na hora que a gente publica. A gente paga e aí ele deixa disponível para tudo, né? Para todas as publicações. Então, é como eu estou te falando, o acesso à ciência é pelas revistas que estão detentoras, eu posso dizer dos dados, das informações. Que a gente transfere os dados para eles de forma fechada, né? Então, só se a gente tiver algum tipo de acesso, que no caso vem pelo próprio portal dos periódicos.

Semelhantemente, PQu9 mencionou algumas contradições relacionadas a Ciência aberta, ao falar que esse fenômeno assumiu um viés comercial, uma vez que muitos periódicos na área cobram para publicar conteúdos em acesso aberto. Sob a perspectiva de PQu9, há um desarranjo na produção científica, no qual o autor ou quem o patrocina está incumbido de custear essa abertura, enquanto tais periódicos se tornam os proprietários da informação publicada e ainda podem explorá-la financeiramente conforme seus interesses. Sobre isso, PQu9 explicou:

Na ciência aberta, a gente paga antecipadamente, uma vez que ele [o artigo] foi aceito, a gente paga para que seja aberto. Então, na verdade, é um modelo um pouco desvirtuado em relação ao que foi pensado inicialmente sobre ciência aberta, porque, atualmente, esse modelo parte do princípio que o autor ou alguém que patrocina o autor tem que arcar com os custos dessa abertura, enquanto que a revista, muitas delas não têm custos, mas elas ficam proprietárias da informação [...] então, de qualquer forma, eu não vejo muito como sendo ciência aberta, é aberta porque foi pago para ser aberta, mas não é uma medida, digamos assim, vamos supor, da ONU ou de uma organização internacional, organização científica que diz assim: ‘agora a ciência vai passar pelo crivo dos pares, mas vai permanecer aberta, não vai ser preciso pagamento para acesso a essa informação’. Então, me parece que o que a gente chama hoje de ciência aberta não é bem uma ciência aberta, ela é aberta a partir do momento que eu garanto que foi feito um pagamento para que isso seja disponibilizado.

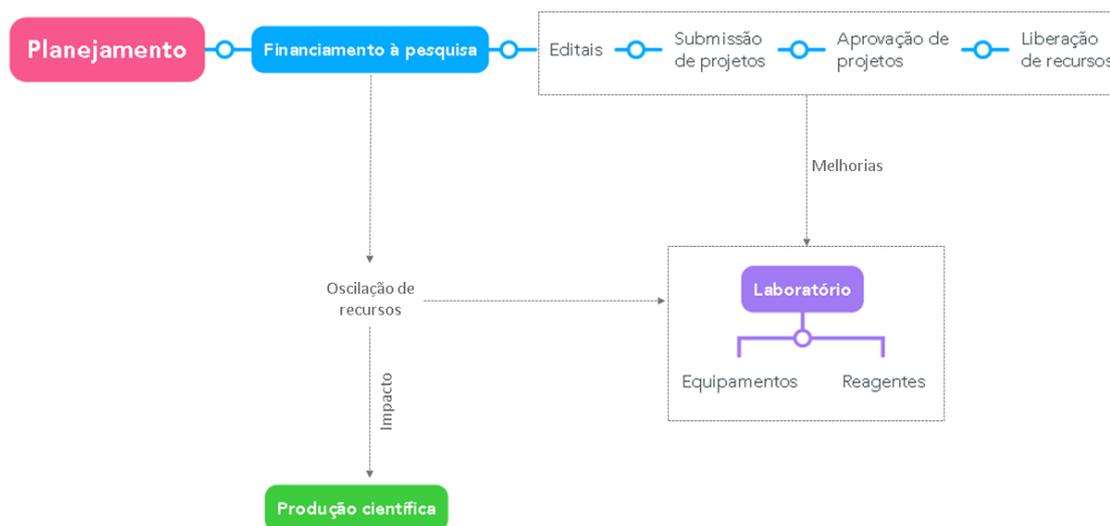
À luz desse contexto, PQu10, sugere maior aprofundamento do tema em âmbito nacional e internacional, sobretudo com vistas ao questionamento do atual modelo de publicação científica. Segundo PQu10, nesse modelo, grandes editores se tornaram proprietários da produção científica, e, por essa razão, acabam requerendo o pagamento de taxas para a publicação e acesso a conteúdos. Entretanto, a Ciência aberta é um tópico de discussão ascendente e tem argumentado em prol de o conhecimento estar disponibilizado livremente, sem barreiras financeiras para ser acessado. Nessa direção, PQu8 e PQu10 afirmaram que pesquisas realizadas com recursos públicos também deveriam ser abertas, uma vez que a sociedade é o principal fomentador desses estudos.

4.2.5.2 Categoria “Planejamento”

A categoria “Planejamento” foi desenvolvida nos termos da subcategoria “Financiamento à pesquisa”, em razão da relevância que tal subcategoria representou no contexto de pesquisa vivenciado pelos entrevistados no estudo. Conforme demonstrado na Figura 32, geralmente, o planejamento de pesquisas é realizado com base em recursos obtidos por meio da aprovação de projetos em editais públicos. Após a liberação desses recursos, pesquisadores viabilizam a execução de projetos aprovados e implementam

melhorias nos laboratórios que coordenam. Tais melhorias incluem a manutenção de equipamentos e a aquisição de novos reagentes para experimentos. Porém, nota-se nesse contexto, uma oscilação nas linhas de financiamento disponibilizadas por agências de fomento, a qual tem impactado negativamente a produção científica na área.

Figura 32 – Categoria Planejamento - Química



Fonte: elaborado pela autora

- Subcategoria “Financiamento à pesquisa”

Com base nos relatos obtidos no estudo, principalmente de PQu1, PQu2, PQu6, PQu7 e PQu9, observou-se que o planejamento de pesquisas, geralmente, é decorrente da participação de pesquisadores em editais de financiamento à pesquisa promovidos por instituições públicas. Apesar da predominância dessas instituições, convém destacar que há também apoio financeiro originário de empresas privadas, todavia, esse é visto em menor proporção.

Com a aprovação do financiamento do projeto de pesquisa, pesquisadores buscam otimizar os recursos recebidos, aplicando-os, quando possível, no melhoramento da estrutura dos laboratórios que coordenam. Conforme destacou PQu2, o custo de pesquisa nesses laboratórios é muito alto, principalmente em razão dos valores necessários à manutenção de aparelhos utilizados em experimentos. Ao se referir a esse aspecto, PQu10 comentou que é fundamental proporcionar condicionamento climático e atmosférico para

a realização de experimentos, principalmente quando se considera as características de um país tropical, com altas temperaturas.

Nesse contexto, pesquisadores como PQu2, PQu4, PQu5 e PQu10 destacaram dificuldades na realização de pesquisas devido à diminuição, cada vez mais frequente, de verbas disponibilizadas por agências de fomento. Sobre tal aspecto, PQu4 comentou que é oneroso manter laboratórios, sobretudo, em razão da manutenção de equipamentos e a aquisição de reagentes:

O dinheiro que as agências de fomento têm trazido para gente, muitas vezes não faz nem a manutenção do equipamento, é só mesmo o custeio e equipamento e ainda tudo limitado [...] então, a gente está voando bem baixinho, entendeu? A nossa pesquisa, ainda mais que na área de ciência, ciência acaba sendo caro. Nossa! Tem reagentes muito caros, que é tudo importado.

Considerando as dificuldades expostas pelos pesquisadores entrevistados, PQu5 reiterou que projetos de pesquisa têm sofrido redução de verbas reservadas para a sua consecução. Por essa razão, é de suma importância que pesquisadores estimem os gastos necessários à pesquisa, prevendo também a alteração de custos, devido a fatores externos. A título de ilustração, PQu5 disse que custos de alguns de seus projetos foram aumentados em consequência da pandemia da Covid, ocasionando variações cambiais na importação de reagentes químicos utilizados em experimentos. Sobre isso, PQu5 explicou:

O dinheiro, ele é muito escasso, ele tem que ser usado de maneira sábia [...] eu tive projetos submetidos antes da pandemia. O recurso foi liberado dentro do período da pandemia. Eu tinha feito uma projeção de gastos para reagentes que ficou inviável, porque o dólar estourou. A luva, a cada 25 reais uma caixa foi para R\$100,00 entendeu? Então, faltou um reagente que custava, sei lá, de R\$ 300, 00 foi para R\$ 600,00, tudo por conta da variação cambial, por conta da pandemia que deixou o mundo maluco. Mas foi um erro meu? Não! Foi um fator de mercado, né? E isso a gente tem que considerar também. Em química, a gente trabalha com muita coisa importada, então, a gente depende dessas variações cambiais. Então, quando a gente manda um projeto, às vezes o dólar está lá em R\$ 5,00, quando o projeto é aprovado, o dólar está R\$ 5,50.

Sob ótica comparativa entre o Brasil e o cenário de produção científica internacional, PQu10 comentou que percebe uma diminuição contínua de recursos destinados à pesquisa, o que prejudica a produção brasileira nesse sentido. Segundo PQu10, a oscilação nas linhas de financiamento no Brasil impacta negativamente o desenvolvimento de novos projetos de pesquisa, além da manutenção da estrutura de pesquisa já existente.

Relacionado a esse contexto, convém apontar que apenas um pesquisador afirmou ter experiência na elaboração de um Plano de gestão de dados. Isso, em razão de atendimento a uma agência de fomento à pesquisa que solicitou tal documento.

4.2.5.3 Categoria “Coleta de dados”

A categoria “Coleta de dados” pode ser entendida conforme suas relações estabelecidas com a subcategoria “Caderno de laboratório”. Assim como é representado na Figura 33, o caderno de laboratório destaca-se nesse contexto como o elemento central na produção de dados, possibilitando o registro da pesquisa, a descrição de processos realizados e a validação de dados por meio de sistematização rigorosa de métodos utilizados.

Figura 33 – Categoria Coleta de dados - Química



Fonte: elaborado pela autora

- Subcategoria “Caderno de laboratório”

Os dados obtidos no estudo demonstraram que apesar da coleta de dados por pesquisadores abranger diferentes técnicas para a produção de dados, como a busca em bases de dados, recolhimento de materiais biológicos e a realização de experimentos, um elemento foi destacado nesse contexto: o caderno de laboratório. Seja em meio físico ou digital, esse instrumento abarca os principais registros da pesquisa científica e do modo pelo qual o conhecimento produzido se materializa no âmbito do laboratório.

Em geral, o caderno de laboratório diz respeito aos processos realizados em experimentos, logo, está diretamente associado à rotina de pesquisadores. Ao falar sobre esse importante instrumento, PQu1 mencionou que costuma utilizá-lo em meio impresso,

mas hoje, cada vez mais, lança mão de cadernos armazenados em nuvem, como se pode ver em sua fala:

Basicamente, temos dois tipos de dados principais: um seria, digamos, o caderno de laboratório, como efetuamos o experimento. Isso, às vezes é abordado de forma física nos cadernos físicos de papel, mas hoje em dia, estamos guardando muito já em formato digital, na maioria dos casos em documentos no Google drive ou planilhas do Google drive.

Reiterando o exposto, PQu10 afirmou que o caderno de laboratório é tão importante para o desenvolvimento de pesquisas, que o mesmo só sai do laboratório o qual coordena, por razões excepcionais. Sendo assim, pode-se notar que não se trata de mero instrumento utilizado por pesquisadores na coleta de dados, mas sim, de um registro que documenta a realização da pesquisa, por meio da descrição de processos utilizados, e de um recurso utilizado para a validação de dados, principalmente relacionados aos procedimentos metodológicos da pesquisa.

A preocupação com o registro de dados em cadernos de laboratório é igualmente percebida em relação à sistematização de métodos adotados pelos entrevistados da área. PQu5 mencionou que os dados que produz passam por rigorosa análise estatística, sendo necessário repetir determinados procedimentos incontáveis vezes, até o ponto de fazer uma inferência relativa ao fenômeno o qual tem interesse de comprovar.

No que se refere à sistematização de procedimentos de pesquisas, PQu4 destacou sua atenção para obter medidas precisas de objetos estudados e que se enquadrem na margem de erro aceitável, conforme apontado na literatura científica da área. Como comentou PQu4, trata-se da obtenção de dados em ambientes constantemente monitorados pelo pesquisador:

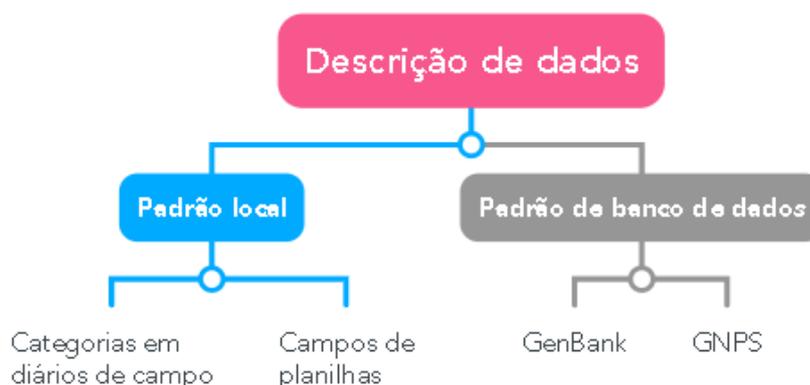
A coleta de dados tem que ser sistemática porque você tem o rigor. Na área de exatas, a gente tem erros que são aceitos e também fala da confiabilidade deles [...] então, o erro é controlado de forma que toda vez que extrapola aquele erro, eu tenho que avaliar ele estaticamente para poder apresentar. Então, todos os nossos dados são rigorosamente controlados e avaliados constantemente, de forma que ele não fuja a um padrão, porque se ele fugir do padrão, quando eu vou para a literatura, a literatura [diz]: ‘olha, o erro é até 20%, você está acima de erro 20%, porque está com esse erro de 20%? Então, tem que discorrer em cima disso, entendeu? Então, o rigor de nossos dados na hora da coleta é do próprio analista, é individual e ele é avaliado, né?

4.2.5.4 Categoria “Descrição de dados”

Conforme demonstrado na Figura 34, a categoria “Descrição de dados” pode ser entendida a partir da subcategoria “Padrão local”. Os dados obtidos no estudo mostraram

que apenas um pesquisador utilizou metadados padronizados para a descrição de conjuntos de dados em duas plataformas internacionais – GenBank e GNPS. Ou seja, os outros pesquisadores afirmaram ser os responsáveis pela descrição dos dados que manuseiam, delineando-se, assim, um padrão local que lança mão principalmente de categorias descritivas e preenchimento de campos de planilhas eletrônicas.

Figura 34 – Categoria Descrição de dados - Química



Fonte: elaborado pela autora

- Subcategoria “Padrão local”

A partir dos dados coletados no estudo, percebeu-se que, em geral, os dados produzidos pelos pesquisadores entrevistados são descritos por eles mesmos, não sendo adotados padrões de metadados direcionados à área de Química ou qualquer outro padrão genérico. Apenas um pesquisador afirmou ter lançado mão de metadados padronizados na submissão de dados para plataformas internacionalmente conhecidas na área.

Sendo assim, dados produzidos pelos entrevistados costumam ser descritos no âmbito do próprio laboratório de quem os gerou. PQu9 ressaltou que apesar de haver normas internacionalmente conhecidas que regulem os procedimentos a serem realizados na coleta de dados, não utiliza metadados para fim de descrição desses.

Ao falar de sua vivência na área da pesquisa, especificamente na manipulação de dados espectrométricos, PQu8 destacou a importância da uniformização de metadados relacionados a esse tipo de dado. De acordo com PQu8, um padrão de metadados aplicado

a dados espectrométricos viabilizaria o reuso de dados e as consequentes comparações entre estudos diferentes, conforme explicou:

Posso dizer que hoje em dia não tem um padrão de coleta de dados, de metadados, digamos assim. Então, nesse sentido, não existe um padrão mundial [...] eu publiquei um artigo onde a gente colocava exatamente isso, que existe uma necessidade de uma uniformização da coleta de dados espectrométricos, de metadados espectrométricos, para que depois, ele possa ser utilizado por qualquer pessoa no mundo todo ou fazer comparações, esse tipo de coisa. Mas hoje não se tem isso ainda não, nós estamos propondo que isso aconteça de fato.

PQu6 disse que já compartilhou dados em duas plataformas internacionais que reúnem conjuntos de dados de diversos países sobre sequenciamento genético e descrição de material molecular, sendo elas, respectivamente: *GenBank* e *Global Natural Products Social Molecular Networking* (GNPS). A descrição de dados é realizada conforme o conjunto de metadados estabelecidos por cada uma dessas plataformas. Sobre sua experiência, PQu6 comentou:

Eu vou te dar dois exemplos de dados que são de domínio público. Por exemplo, a gente trabalha com fungos e fungo para ser identificado, é através de análise de biologia molecular, sequenciamento de genes. A gente gera uma sequência de genes e vai lá no repositório que é chamado de GenBank, que é de genes, da NCIBI, que é a National Center for Biotechnology Information. É um repositório de domínio público do mundo inteiro. Você vai lá e deposita a sua sequência, então, fica aberto para quem isolar um fungo similar poder fazer o comparativo, as similaridades [...] e [a gente] faz o uso também de uma outra plataforma que é chamada GNPS, Global Natural Product Social Molecular Networking, também de domínio público. Essa bem mais recente, é de 2017 para cá, e a gente vai lá e deposita os nossos dados brutos e aí outros pesquisadores do mundo inteiro podem acessar esses dados e ver que a gente identificou um determinado metabólico com base naqueles espectros.

De acordo com a *National Library of Medicine*⁷² o GenBank é um banco de dados localizado no *National Institutes of Health* (NIH), que armazena e veicula sequências de ácido desoxirribonucleico (DNA). Além disso, o GenBank integra a rede *International Nucleotide Sequence Database Collaboration*, juntamente com o *DNA DataBank of Japan* (DDBJ) e o *European Nucleotide Archive* (ENA), havendo intercâmbio de dados e informações diariamente. Destaca-se, ainda, que o GenBank tem o objetivo de estimular o acesso da comunidade científica a dados atualizados sobre sequenciamento genético. Nessa direção, não estabelece restrições de uso ou compartilhamento dos dados que veicula.

Segundo informações disponibilizadas no site da GNPS⁷³, esse banco de dados pode ser entendido da seguinte maneira:

É um ecossistema de espectrometria de massa baseado na web que visa ser uma base de conhecimento em acesso aberto para a organização e compartilhamento da ampla comunidade de dados de espectrometria de

⁷² Disponível em: <https://www.ncbi.nlm.nih.gov/genbank/>

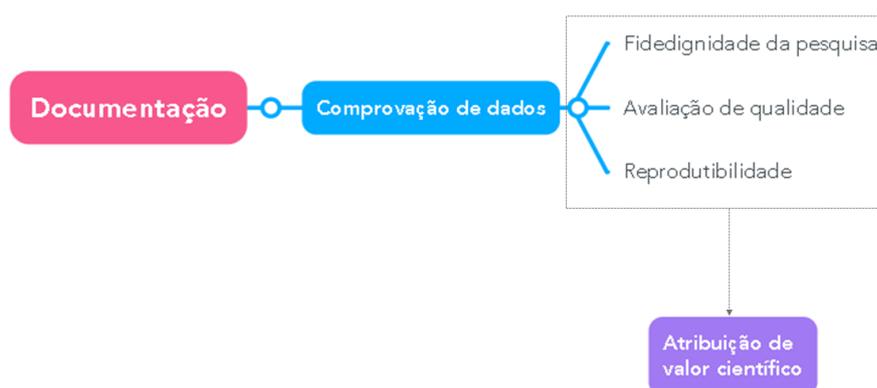
⁷³ Disponível em: <https://gnps.ucsd.edu/ProteoSAFe/static/gnps-splash.jsp>

massa brutas, processadas, ou fragmentos de sua identificação. O GNPS auxilia na identificação e descoberta ao longo de todo o ciclo de vida dos dados; desde a aquisição/análise inicial dos dados até a publicação posterior.

4.2.5.5 Categoria “Documentação de dados”

Conforme mostra a Figura 35, a categoria “Documentação de dados” pode ser compreendida nos termos da subcategoria: “Comprovação de dados”, que, por seu turno, refere-se à verificação de registros relacionados à execução de pesquisas. Nesse sentido, tais registros viabilizam a verificação da fidedignidade de estudos, a avaliação da qualidade de dados e a reprodutibilidade de processos, sendo possível, dessa maneira, atribuir valor científico ao conhecimento produzido.

Figura 35 – Categoria Documentação de dados - Química



Fonte: elaborado pela autora

- Subcategoria “Comprovação de dados”

A partir dos dados coletados no estudo, principalmente nos relatos de PQu1, PQu4, PQu5, PQu8, PQu9 e PQu10, pode-se observar que a documentação relacionada a processos inerentes à pesquisa é fundamental para a comprovação de dados, e, por conseguinte, para atestar que determinado estudo foi realizado conforme parâmetros sistemáticos. Nessa direção, PQu5 chegou a afirmar que se não houver documentação como maneira de disciplinar os processos envolvidos nas pesquisas que realiza, as mesmas não terão valor científico algum.

PQu8 comentou que a documentação é relevante para a comprovação da

fidedignidade da obtenção de dados na área, tal como enfatizou:

Isso [a documentação] é fundamental, é essencial, se não tem isso, você não garante que aquele dado é um dado fidedigno, você não pode garantir que aquele resultado realmente representa alguma coisa. Sem isso, você pode chegar lá no final do resultado e descobrir a cura do câncer, mas se isso não foi feito previamente com todos os cuidados, essa cura do câncer pode ser totalmente uma coisa, assim, equivocada. Por exemplo, isso é essencial, é uma questão chave.

Associada à comprovação de dados, a documentação também é fonte para avaliação do nível de confiança sobre a coleta de conjuntos de dados. Trata-se de mais um recurso utilizado por pesquisadores para avaliarem a qualidade desses. Logo, PQu10 se autodescreveu como sendo um experimentalista, sendo imprescindível trabalhar com dados confiáveis, conforme mencionou:

Fundamental, a gente faz isso o tempo todo, senão já era. Se você não tem qualidade, eu costumo dizer o seguinte: eu preciso de duas coisas para fazer uma boa pesquisa. Primeiro, é ideias, é a pergunta que eu quero responder e como é que eu vou responder essa pergunta. Dois, a qualidade, a confiança da medida do dado, se eu não tenho confiança no dado, não tem pesquisa, eu sou experimentalista [...] a gente faz dez, vinte, cinquenta testes por dia, aí você consegue trabalhar. Então é isso, toda técnica a gente tem um padrão que a gente vai repetindo mês a mês para entender.

Um outro aspecto relacionado à documentação sobre dados de pesquisa e sua comprovação científica é a reprodutibilidade de procedimentos metodológicos descritos e a consequente obtenção resultados idênticos. Conforme mencionou PQu1, registrar o passo a passo da pesquisa é uma boa prática no âmbito da produção científica, principalmente, por viabilizar a reprodução de experimentos, ou seja, do conjunto de protocolos executados no contexto do laboratório. Sendo assim, convém destacar fala de PQu1 que abordou o aspecto em questão:

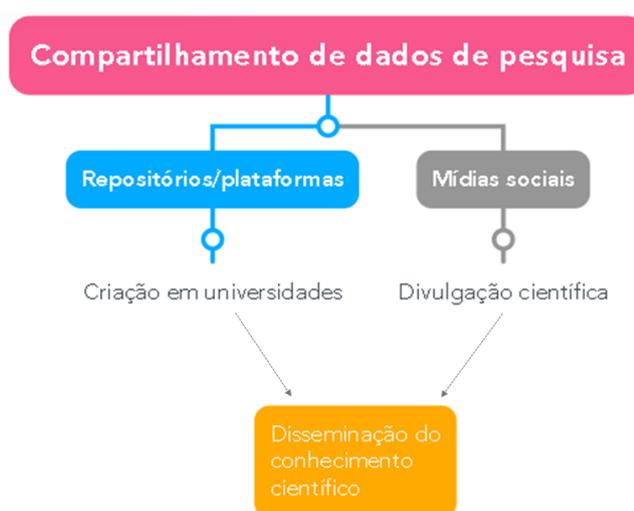
Nesse caso, digamos que tentamos seguir o que chamaremos: as boas práticas de laboratório, de que o que você escreve, o que você fala no laboratório, o que você publica permita reprodução dos experimentos. Nesse aspecto, eu tento sempre que todos os protocolos que nós publicamos permitam a reprodução da pesquisa. [...] sempre tentamos dar tudo aquilo que é essencial para outros grupos de pesquisas poderem continuar essa pesquisa ou poderem produzir. Então, tentamos um pouco o que seriam as boas práticas em química.

4.2.5.6 Categoria “Compartilhamento de dados”

A categoria “Compartilhamento de dados” foi desenvolvida a partir da subcategoria “Repositórios” (que também abrange outras plataformas de disponibilização de dados de pesquisa), conforme é apresentado na Figura 36. Alguns entrevistados relataram já ter compartilhado seus dados nos referidos repositórios e plataformas, assim como afirmaram ter disponibilizado dados por meio de mídias sociais, incluindo *You tube* e *Instagram*.

Entretanto, percebe-se que repositórios institucionais de dados abertos de pesquisa ainda se encontram em processo de criação e desenvolvimento nas universidades da maioria dos entrevistados, apesar de caracterizada a importância dessas plataformas no contexto de disseminação do conhecimento científico.

Figura 36 – Categoria Compartilhamento de dados - Química



Fonte: elaborado pela autora

- Subcategoria “Repositórios/ plataformas de dados abertos”

A partir das informações coletadas no estudo, pode-se observar que o compartilhamento de dados por meio de repositórios ou outras plataformas abertas semelhantes ainda não é uma realidade comumente vista pelos pesquisadores entrevistados. Apenas PQu6 e PQu8 afirmaram ter compartilhados dados por meio desses canais. Já PQu1 e PQu5, comentaram que há projetos de implementação de repositório de dados de pesquisa nas universidades em que atuam, mas que esses ainda se encontram em fase de discussão inicial.

PQu8 expôs sua experiência no compartilhamento de dados abertos em uma plataforma criada pela universidade a qual integra. Trata-se de pesquisa que lançou mão de amostras de sangue de pacientes para o desenvolvimento de um método eficaz para o diagnóstico de doença encefálica. Sendo assim, PQu8 compartilhou publicamente dados originais da pesquisa realizada, bem como dados relativos aos equipamentos utilizados no

estudo, na plataforma mencionada. Sob seu ponto de vista, tal medida foi importante para que outros pesquisadores pudessem reutilizar os dados compartilhados, avaliando-os novamente, ou acrescentando-lhes outras informações. A respeito disso, PQu8 comentou:

Aqui na universidade, o pessoal da informática, eles têm lá uma plataforma onde você pode disponibilizar os seus dados brutos, chama-se de dados brutos ou dados originais. Então, a gente colocou esses dados originais, que estão disponíveis lá. Então, agora qualquer pessoa de qualquer lugar do mundo pode pegar esses mesmos dados e avaliar de novo, até refazer o estudo que a gente fez para ver se realmente está certo ou ver se acha mais informações.

Por sua vez, PQu6 afirmou ter compartilhado dados de pesquisa sobre sequenciamento genético e descrição de material molecular em dois bancos de dados internacionais chamados, respectivamente: *GenBank* e *Global Natural Products Social Molecular Networking* (GNPS). Para PQu6, tal compartilhamento possibilitou que qualquer outro pesquisador pudesse analisar diferenças e semelhanças de genes.

Com relação à criação e desenvolvimento de repositórios de dados de pesquisa no ambiente de universidades, PQu1 destacou que a universidade a qual integra possui um projeto de criação de repositório de dados. Segundo PQu1, o referido repositório conterá metadados atribuídos a conjunto de dados e apresentará relações do DOI com as publicações correspondentes.

A universidade em que atua PQu5 também está em processo de criação de um repositório de dados. Os técnicos responsáveis pelo projeto estão desenvolvendo estratégias de filtros na pesquisa sobre dados, com a finalidade de proteger alguns desses que não poderiam ser expostos sem a definição de um nível de segurança mínimo para serem acessados. Enquanto isso, bem como para divulgar a ciência em linguagem mais acessível à sociedade, PQu5 lança mão de mídias sociais e plataformas como um canal no *you tube* e uma conta no *Instagram*. Em linhas gerais, trata-se de estratégias criadas para publicizar o conhecimento produzido no ambiente de universidades. A fala de PQu5 ilustra os aspectos mencionados nesse sentido:

Até o momento, nós não tivemos essa experiência, como eu te mencionei anteriormente, a possibilidade de desse repositório, né? Dentro da universidade seria de forma aberta, mas é uma ideia que está tentando ser construída, porque a equipe da computação com o pessoal da matemática está tentando desenhar a melhor forma de a gente abrir esses dados, e aí seria pesquisa global, mas da mesma maneira, a gente tem que ter alguns filtros de acessos para que essas informações não sejam usadas de maneira errada. Por que tudo na vida é uma estrada de mão dupla, se você liberar demais, você incorre, às vezes, dessa informação ser usada de maneira inadequada. Mas até o momento, é o único tipo de informação. A gente, dentro do nosso grupo de pesquisa, a gente tem um canal no youtube e tem um canal no instagram. Então, o que a gente tem feito? Vídeos dentro de uma linguagem mais acessível para as pessoas entenderem: 'olha, como que eles estão

trabalhando?’ Claro que as vezes você não pode fugir de um termo técnico ou outro, mas é tentar fazer vídeos de dois, três minutos, no máximo, que expliquem quais são as pesquisas, o que os alunos fazem, quais são os projetos deles. E aí, tem tido uma boa aceitação.

4.2.5.7 Categoria “Preservação de dados”

Conforme representado na Figura 37, a compreensão da categoria “Preservação de dados” ocorre em função das relações estabelecidas com a subcategoria “Estratégias locais”. Isso porque, geralmente, pesquisadores lançam mão de procedimentos individuais com o objetivo de preservar os dados que manipulam. As iniciativas institucionais nesse sentido demonstram-se timidamente e revelam vulnerabilidades relativas à preservação de dados a longo prazo na área.

Figura 37 – Categoria Preservação de dados - Química



Fonte: elaborado pela autora

- Subcategoria “Estratégias locais”

As informações obtidas nas entrevistas realizadas demonstraram que a preservação de dados ocorre mediante à adoção de estratégias locais, ou seja, trata-se de uma preocupação predominantemente individual, e não institucional. Sendo assim, as principais medidas adotadas pelos pesquisadores podem ser resumidas da seguinte maneira:

- armazenamento de dados em nuvem;
- armazenamento de dados em discos rígidos;
- armazenamento de dados em computadores pessoais ou pertencentes à universidade na qual o pesquisador atua;
- envio de dados por e-mails;
- assinatura de termo de sigilo e confidencialidade de cada membro da equipe de pesquisa, com o objetivo de prevenir a perda do ineditismo ocasionado por uma publicação de dados equivocada;

- registro de procedimentos experimentais em cadernos de laboratório impressos ou digitais;
- manutenção de equipamentos e computadores, o que inclui o condicionamento climático necessário para a preservação de resultados de pesquisas, e, por conseguinte, de dados obtidos.

A fala de PQu6 ilustra alguns dos procedimentos de preservação mencionados logo acima:

Então a gente vai gerando esses dados, os alunos vão obtendo os dados de espectrometria de massas, eles vão compartilhando comigo, eu vou discutindo com os alunos, mas todos os arquivos ficam armazenados. A gente peca pelo excesso, a gente armazena em HD, o meu HD do meu notebook, no computador daqui da [universidade em que atua], no computador do aluno, na nuvem, a gente compartilha em drives e no final, como eu falei, no repositório da [universidade em que atua].

Apenas PQu6 afirmou ter experiência na elaboração de um Plano de preservação de dados, sendo que tal documento foi exigido por uma agência de fomento, como cumprimento de requisito para o financiamento de sua pesquisa. Conforme destacaram PQu7 e PQu8, a preservação de dados é realizada predominantemente em nível individual, não havendo instrumentos institucionais amplamente adotados nesse sentido.

Conseqüentemente, tal situação evidencia a fragilidade relacionada à proteção e preservação de dados a longo prazo na área, de modo que não há garantias de que os mesmos estarão disponíveis e serão acessados posteriormente. Logo, há riscos de perda de dados e de descontinuidade de leitura desses, devido à obsolescência de mídias para a sua reprodução.

4.2.5.8 Categoria “Reuso de dados”

Como pode ser observado na Figura 38, a categoria “Reuso de dados” foi desenvolvida nos termos da subcategoria “Trabalhos publicados”. Isso porque, verificou-se dentre a maioria dos relatos obtidos nas entrevistas, que é comum a reutilização de informações publicadas no âmbito da literatura científica, e não de dados brutos ou ainda não publicados. Sendo assim, geralmente, pesquisadores citam em seus textos trabalhos publicados, o que lhes proporciona fazerem comparações entre resultados e confrontarem dados coletados. Nesse contexto, o reuso de dados, sem a devida responsabilização

autoral, é eticamente reprovável. Convém destacar, ainda, que dados brutos são reutilizados, especialmente, para darem continuidade a estudos já realizados por pesquisadores que os produziram.

Figura 38 – Categoria Reuso de dados - Química



Fonte: elaborado pela autora

4.2.5.9 Trabalhos publicados

Os dados coletados neste trabalho demonstraram que, dentre os entrevistados, o reuso de dados brutos, ou ainda não publicados, não é comum na área. Isso porque predomina-se o reuso de informações publicadas, as quais devem ser devidamente citadas quando forem inseridas em novos textos. Com base nos relatos de PQu1, PQu2, PQu4, PQu5, PQu7 e PQu10, observou-se que não houve menção direta à citação de dados, mas sim, a trabalhos publicados e que integram a literatura científica da área.

Para PQu2, o reuso de dados de outro pesquisador sem a devida autorização e atribuição de autoria é uma atitude eticamente reprovável. Nesse sentido, destacou que costuma se valer de outros estudos publicados para reproduzir condições experimentais,

sempre os referenciando adequadamente. Além disso, PQu2 comentou que procura se atualizar em relação à literatura científica produzida na área, o que é muito importante principalmente para obter conhecimento de novos métodos, conforme explicou:

Você usar um dado experimental de um colega como seu, isso é em linhas gerais, eu não lembro de nenhuma exceção, mas isso seria eticamente reprovável, certo? Absolutamente reprovável. O que nós fazemos? Muitas vezes sim, você cita um trabalho de um colega, você cita um trabalho de alguém, dados que ele produziu para embasar, justificar ou mesmo corroborar dados que você também tenha obtido. Então, em uma determinada situação, um colega, um pesquisador encontrou um tal resultado em um determinado experimento, nós estudando a nossa realidade fizemos o mesmo experimento e chegamos a mesma conclusão que ele havia chegado na situação dele, só que agora na nossa, então, você pode sim se valer de citações de dados de outrem para essa finalidade, certo? Uma outra coisa que acontece muito é você utilizar uma condição experimental já validada por outro pesquisador para aplicar nos seus experimentos, isso é extremamente difundido. Aliás, eu acho que essa é umas das principais informações que a gente busca quando faz o acompanhamento da literatura, primeiro, vamos ver o que estão fazendo por aí e como estão fazendo, né? Porque esse como está fazendo é muito importante porque a gente está tendo a todo dia novas publicações surgindo com novos métodos para se fazer a mesma coisa, métodos ambientalmente mais corretos, mais amigáveis, mais baratos, mais ágeis, que são mais fidedignos. Então, você, volta e meia, está reusando informações, tecnologias e metodologias que estão divulgadas no meio científico.

De acordo com PQu10, um pesquisador só sabe se dados coletados têm medidas maiores, menores, melhores ou piores, quando os compara com os dados presentes na literatura científica. Semelhantemente, PQu5 e PQu7 disseram que o reuso de informações de trabalhos publicados é importante para confrontar dados coletados, compará-los, bem como validá-los. Sobre isso, PQu7 fez o seguinte comentário:

Às vezes acontece, por exemplo, de você fazer um trabalho e comparar os teus resultados com o do outro pesquisador. Isso já teve, os meus dados para outros pesquisadores e eu já comparei os meus com os deles. Até mesmo para dizer que estava indo na linha certa, mas daí, tem que citar e cada dado que é mencionado: 'ah, o fulano fez tal coisa, ah, o fulano encontrou tal resultado, nós encontramos esse resultado'. Então, a gente tem muito na parte de química, a parte de comparação.

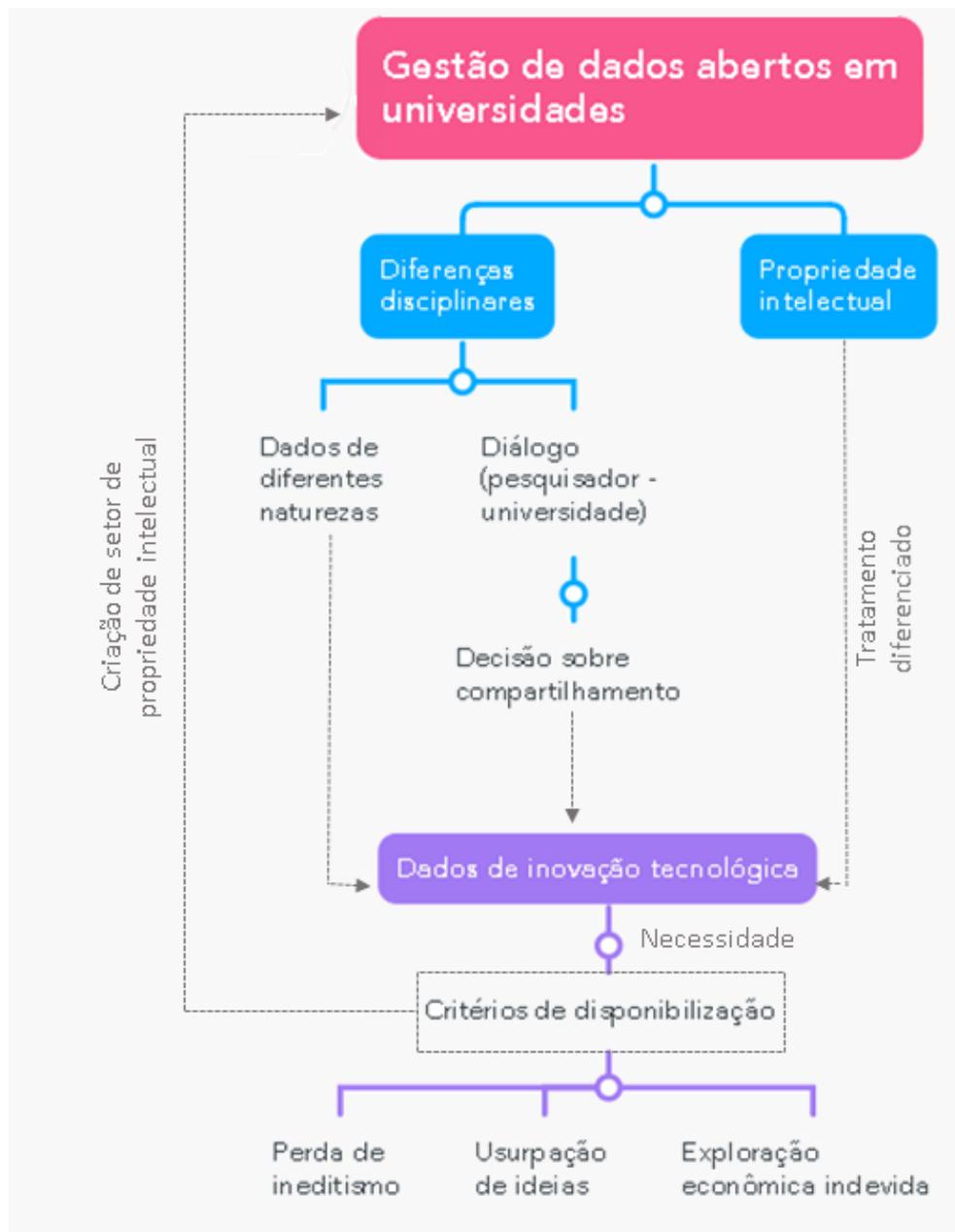
Apesar de a maioria dos entrevistados não terem abordado, especificamente, o reuso de dados brutos ou ainda não publicados, mas sim da própria literatura científica, tais dados são reutilizados, em especial, para dar continuidade a estudos realizados por pesquisadores que os produziram. Como ilustrou PQu6, certa vez, recorreu a dados produzidos na década de 1980 para retificar dados publicados na literatura e atualizá-los conforme a utilização de novas tecnologias aplicadas à pesquisa.

4.2.5.10 Categoria “Gestão de dados abertos de pesquisa em universidades”

Conforme mostrado na Figura 39, a categoria “Gestão de pesquisa em universidades” foi desenvolvida nos termos de duas subcategorias: “Diferenças disciplinares” e “Propriedade intelectual”. Sobre a primeira, destaca-se que há diferentes tipos de dados, tais como aqueles produzidos na área de inovação tecnológica. Acrescente-se, por oportuno, que entrevistados enfatizaram o diálogo entre pesquisadores e universidades como um instrumento de grande valia no processo de tomada de decisão sobre o compartilhamento de dados, principalmente, os tecnológicos.

A subcategoria “Propriedade intelectual” evidencia que alguns dados, como os de natureza tecnológica, necessitam de tratamento diferenciado no que se refere aos critérios de sua disponibilização. Isso com o propósito de minimizar os riscos de perda de ineditismo, usurpação de ideias e exploração econômica indevida. A definição dos critérios mencionados sugere a criação ou desenvolvimento de um setor responsável pela propriedade intelectual em universidades, o qual trataria de aspectos relacionados à gestão de dados de pesquisa nessas instituições.

Figura 39 – Categoria Gestão de dados abertos de pesquisa em universidades – Química



Fonte: elaborado pela autora

- Subcategoria “Diferenças disciplinares”

A partir dos dados obtidos no estudo, pode-se perceber que diferenças disciplinares influenciam a gestão de dados de pesquisa no contexto de universidades. Isso, principalmente, pelo fato de essas instituições reunirem diversas áreas do conhecimento, cada qual com sua maneira de produzir ciência e comunicá-la. De acordo com PQu1, dados

divergem entre si conforme a sua natureza, e o manejo dos mesmos varia conforme o contexto disciplinar em que foram produzidos.

Ao propor alguma sugestão ou recomendação para sua universidade, caso a gestão de dados abertos seja institucionalizada, PQu5 destacou que é fundamental reconhecer as especificidades de cada área do conhecimento. A título de exemplo, citou que pesquisadores que lidam com inovação tecnológica ou com o desenvolvimento de novos produtos poderiam vivenciar alguns entraves relacionados à divulgação de dados. Infere-se de sua fala que caberia à universidade estabelecer as diretrizes ou o conjunto de critérios necessários para tal divulgação, prestar esclarecimentos sobre quando dados poderiam ser publicados (se durante a execução da pesquisa ou após a mesma ter sido concluída) e estabelecer os níveis de permissão de acesso a dados (se total ou parcial):

Eu acho que é salutar, a gente vai ter que aprender a fazer isso de uma maneira muito rápida, provavelmente eu acho que vai ter que ver quais são as especificidades de cada área. Para quem trabalha com tecnologia, por exemplo, com desenvolvimento de produto, como é o nosso caso e algumas linhas, eu acho que alguns entraves devem existir. Eu não sou contra divulgar meu dado bruto, mas em que momento eu vou fazer isso, por exemplo? Após a conclusão da pesquisa que eu tenho o depósito da patente ou durante a execução da pesquisa? Então, particularmente, eu acho que a gente tem que discutir os critérios, quais são os critérios dessa forma de gestão de dados, se eles vão ser disponibilizados concomitante à execução da pesquisa, eles vão ser disponibilizados após a execução da pesquisa com devida proteção, e se esse acesso vai ter algum tipo de filtro. Ou seja, qualquer pessoa da comunidade tem liberação a tudo, ou há diferentes níveis de informação? Sei lá. Se for de uma empresa, será que ela vai ter acesso a tudo? Pode ser que sim, desde que ela se comprometa a não usar aqueles dados, ou nada daquela pesquisa sem a comunicação prévia, ou algo assim. Talvez eu acho que a gente vai ter que discutir efetivamente quais são os critérios.

Semelhantemente, na fala de PQu3 observa-se que caberia à universidade estabelecer canais de comunicação com seus pesquisadores a respeito da gestão de dados. Assim, destaca-se a importância do diálogo entre instituição e pesquisador, de modo que a opinião desse último também seja relevante para determinar a tomada de decisão sobre compartilhar dados, ou não, uma vez que o pesquisador não é obrigado a assim fazê-lo. Mais uma vez, PQu3 ressaltou que o campo de inovação tecnológica deve ser considerado nesse contexto, na medida em que nem todos os dados poderiam ser disponibilizados publicamente.

Eu acho que seria o inverso, né? Eu acho que a instituição é que tem que procurar o pesquisador para procurar saber, tem que ser completamente o inverso disso daí, porque se eu estou gerando um determinado dado, né? Eu não tenho obrigação de compartilhar ele, nenhuma, eu não tenho obrigação nenhuma de compartilhar e a universidade ela tem que ouvir os pesquisadores, tem que haver um consenso. Eu não estou dizendo também que sou contra compartilhar, existem coisas que podem ser compartilhadas e devem ser compartilhadas, mas existem outras coisas que não, quem trabalha com a parte de inovação tecnológica é um problema isso daí.

- Subcategoria “Propriedade intelectual”

Um dos aspectos mais abordados em relação à gestão de dados de pesquisa em universidades é a propriedade intelectual, visto que muitos entrevistados demonstraram preocupação a seu respeito. Nessa direção, PQu2 explicou que em relação à área tecnológica, há interesses econômicos que permeiam a exposição de dados. Isso quer dizer que em um contexto altamente competitivo e inovativo, aqueles que têm acesso a dados e informações privilegiadas obtêm valor agregado e podem alcançar algum tipo de vantagem, como exploração econômica de determinado composto químico. Por essa razão, PQu2 afirmou que no âmbito das ciências duras, especialmente, há dados que necessitam estar sob sigilo:

Eu acho que eu sou muito a favor de que tudo que puder ser divulgado, transparente seja! Mas é evidente que quando você vai para a área tecnológica, digamos assim, as ciências mais duras, que não é das áreas humanas, você começa a ter toda a questão de interesse econômico e você, invariavelmente, você não pode ter uma exigência de divulgação de resultados sob o ponto de você ter comprometimento da questão tecnológica, da questão inovativa, da competição, da geração de conhecimento de valor agregado, você vai estar mexendo na soberania nacional, você vai estar mexendo na economia. Então, isso tem que ser muito bem tratado, né? Tem coisas que merecem ser escondidas para dar dinheiro, outras merecem ser divulgadas com a mesma finalidade, temos que tratar cada caso com a sua particularidade.

Outra preocupação relacionada à propriedade intelectual e que foi demonstrada pelos entrevistados é a perda do ineditismo de dados, ocasionada pela publicação antecipada desses por outros pesquisadores. Mais uma vez, reiterou-se que áreas de produção tecnológica são suscetíveis à usurpação de autoria de ideias no contexto de uma mesma universidade, de universidades geograficamente próximas, ou distantes. Sobre isso, PQu3 declarou:

Quem trabalha com a parte de inovação tecnológica é um problema, isso daí. Eu trabalho com uma parte de inovação tecnológica, então eu vou ficar publicando? Não! E se eu publico aqui, o que vai acontecer amanhã? Lá na Europa aparece um cara fazendo a mesma coisa, entendeu? Isso acontece muito aqui dentro da própria instituição e nas instituições vizinhas. Eu trabalhar com determinada coisa e quando ver, a pessoa está publicando material com aquilo que eu estava trabalhando, claro, você fica assim: meu Deus!

Nesse sentido, PQu5 enfatizou a importância de se garantir ao pesquisador a proteção de dados produzidos, visando mitigar riscos relacionados à apropriação indevida e à publicação antecipada de dados por indivíduos não autorizados. Para PQu5, como o ritmo de produção no ambiente acadêmico é diferente do setor produtivo, é preciso adotar

mecanismos de proteção a dados, tais como patentes, antes da abertura dos mesmos. A esse respeito, PQu5 comentou:

A velocidade de produção é outra da universidade em relação ao setor produtivo, você sabe muito bem disso, se eu coloco isso na plataforma aberta e não tiver protegido, não tiver ao mínimo uma solicitação de patente, qualquer empresa pode pegar e mudar um pouco a ideia e adiantar, né?

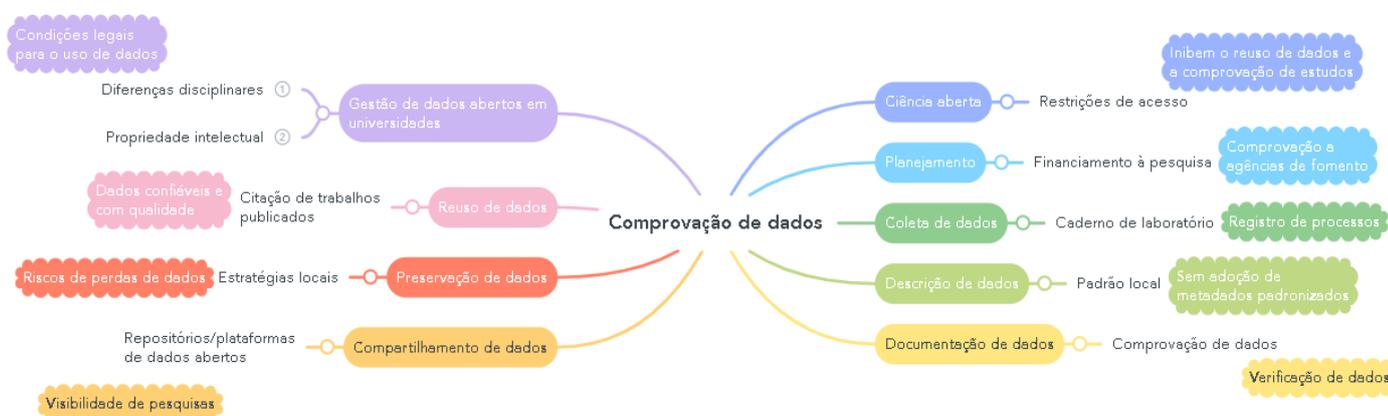
Destaca-se também as comparações feitas por PQu7 sobre medidas de proteção de propriedade intelectual realizadas no Brasil e em países desenvolvidos, tais como Estados Unidos. Segundo PQu7, o Brasil concentra-se na concessão de patentes em nível de produtos, enquanto outros países já liberam patentes em nível de ideias.

À luz desse contexto, faz-se relevante mencionar perspectiva de PQu4, na qual sugere-se a criação ou desenvolvimento de setores em universidades que sejam responsáveis pela propriedade intelectual produzida por pesquisadores, bem como pela gestão da abertura de dados.

4.2.6 Codificação seletiva e teoria substantiva relacionada à gestão de dados de pesquisa na área de Química

No que se refere à área da Química, dados obtidos no estudo constataram a centralidade da categoria “comprovação de dados” nas relações estabelecidas com as demais categorias identificadas. Assim como é representado na Figura 40, pode-se destacar tais relações da seguinte maneira:

Figura 40 – Comprovação de dados e sua relação com práticas de gestão de dados de pesquisa – Química



Fonte – Elaborado pela autora

- a exigência do pagamento para a publicação de dados abertos restringe o acesso a conteúdos científicos, principalmente pelo fato de pesquisadores não disporem de recursos financeiros suficientes para o custeio desse tipo de publicação. Tais restrições obstaculizam o reuso de dados que poderiam ser insumos para a comprovação de estudos realizados;
- o planejamento relacionado a dados de pesquisa no campo da Química, geralmente, está atrelado a recursos financeiros recebidos por cientistas para a execução de pesquisas. Nesse sentido, dados assumem um caráter comprobatório em frente a agências de fomento, na medida em que são registradas descobertas obtidas no decorrer do projeto de pesquisa;
- no que se refere à coleta de dados de pesquisa na área de Química, observa-se o lugar de destaque ocupado por cadernos de laboratório, principalmente, por esses

registrarem procedimentos de pesquisa, servindo também de fonte para a comprovação de dados obtidos;

- pesquisadores na área da Química costumam adotar seus próprios procedimentos para descrever os conjuntos de dados que produzem, sem a utilização de metadados, os quais poderiam facilitar a recuperação e posterior comprovação de resultados de pesquisas;
- em geral, a documentação de procedimentos relacionados a pesquisas na área de Química está relacionada à verificação de dados para fins comprobatórios;
- repositórios de dados podem maximizar a visibilidade de pesquisas, sendo um canal para recuperá-los, analisá-los e comprová-los quanto às informações que veiculam;
- em geral, pesquisadores de Química adotam suas próprias estratégias para a preservação dos dados que geram, implicando em riscos que comprometem o reuso de dados a longo prazo. Conseqüentemente, não há garantias efetivas de que dados permanecerão disponíveis para consulta, análise e posterior checagem, a fim de serem comprovados;
- sendo evidenciado que pesquisadores da área de Química prezam por dados confiáveis e de qualidade comprovada cientificamente, constatou-se que muitos cientistas reutilizam informações provenientes de trabalhos publicados na literatura, e não de dados brutos;
- é fundamental que a gestão de dados de pesquisa no contexto universitário considere a influência de diferenças disciplinares. Na área da Química, por exemplo, o aspecto da propriedade intelectual em relação a dados de pesquisa é algo extremamente relevante, principalmente, no que diz respeito a elementos que comprovem a realização de uma pesquisa científica e as condições legais para o seu reuso em outros trabalhos.

À luz da literatura, a próxima seção abordará a categoria central – comprovação de dados de pesquisa – de acordo com as relações estabelecidas com as demais categorias identificadas no estudo, que, por seu turno, dizem respeito às práticas de gestão de dados de pesquisa.

4.2.6.1 Comprovação de dados de pesquisa e sua relação com práticas de gestão de dados de pesquisa

Neste trabalho, observou-se que há restrições de acesso a conteúdos científicos, principalmente, devido ao pagamento de taxas de publicação exigido por editores a pesquisadores. Isso ocorre até quando a publicação é realizada em periódicos em acesso aberto, por exemplo. Por conseguinte, muitos autores brasileiros não dispõem de recursos financeiros suficientes para custear esse tipo de publicação, que muitas vezes, atinge a escala de milhares de dólares no caso de alguns periódicos internacionais.

Com base no exposto, destaca-se que muitas publicações possuem um viés comercial resultante de tratativas estabelecidas entre autores e editores científicos. De modo geral, tal situação implica em restrições de acesso a dados que poderiam ser reutilizados em outros estudos, ou até mesmo ser apreciados por outros cientistas, de modo a ampliar a comprovação da veracidade de informações veiculadas por meio de dados de pesquisa.

Em perspectiva contrária à mencionada, o qual destaca-se a cobrança para a publicação de conteúdos científicos em acesso aberto, Jomier (2017) afirma que a Ciência aberta contribui para a aceleração do processo de publicação científica, contribuindo para que o maior número de informações seja compartilhado. Na visão da autora, um dos pilares da Ciência aberta é o acesso aberto. Apesar da importância da publicação nesses moldes, ainda se nota que alguns periódicos disponibilizam gratuitamente conteúdos científicos apenas após determinado período de embargo. Por conseguinte, o acesso a conteúdos mais recentes é prejudicado, pois geralmente deve ser obtido por meio de compra.

Críticas ao modelo comercial proposto por editores na produção e comunicação científica não são atuais. Segundo Mueller (2006), periódicos revisados por pares são os canais preferenciais para a publicação de artigos, além de serem responsáveis pela chancela do conhecimento científico. Entretanto, esse mérito não foi suficiente para desestabilizar o tradicional sistema de comunicação científica em âmbito global, na década de 1980, após a ocorrência da denominada crise dos periódicos.

Dentre os motivos que ocasionaram a crise dos periódicos, podem-se ressaltar as dificuldades de bibliotecas acadêmicas e centros de pesquisas em atualizar suas coleções de periódicos científicos, e ainda atenderem as demandas informacionais de seus usuários, além de restrições financeiras de editoras decorrentes dos altos custos de produção de periódicos (MUELLER, 2006).

De acordo com Alonso Arévalo, Subirats Coll e Martínez Conde (2008), a crise dos periódicos foi o estopim que abriu espaço para a ascensão do movimento de acesso aberto contra restrições de acesso estabelecidas por editores. Contudo, só foi possível concretizar esse fenômeno devido a utilização de redes de informação, que são canais para o compartilhamento de registros do conhecimento científico.

No que se refere ao planejamento relacionado a dados de pesquisa na área de Química, este estudo constatou que tal prática ocorre principalmente devido à organização financeira de pesquisadores em relação a recursos recebidos para a execução de pesquisas. Considerando a definição da Universidade de Leeds⁷⁴ na qual dados são entendidos como informações coletadas que comprovam resultados originais obtidos em uma pesquisa, nota-se que esses também assumem caráter comprobatório a financiadores de pesquisa, de modo a registrar descobertas alcançadas durante o ciclo de um projeto específico.

Tendo como base o exposto, demonstra-se a importância da disponibilização de dados abertos de pesquisa a instituições financiadoras de projetos. A relevância mencionada não se restringe a esse contexto, uma vez que dados abertos poderiam ser amplamente compartilhados e, por conseguinte, desencadear novos ciclos da produção científica. Contudo, a abertura de dados não deve ocorrer de modo negligente e sem planejamento.

De acordo com a Agência para a Modernização Administrativa (AMA) (2016), há cinco etapas a serem cumpridas por instituições para que dados sejam efetivamente abertos, as quais podem ser explicadas da seguinte maneira:

74 Disponível em:

https://library.leeds.ac.uk/info/14062/research_data_management/61/research_data_management_explained

- **planejamento** – definição de estratégias relacionadas a disponibilização de dados abertos que incluem a identificação de formatos reutilizáveis, o estabelecimento das condições necessárias à atualização de dados e a responsabilização autoral sobre o compartilhamento de dados;
- **definição dos dados que serão abertos** – escolha dos dados que serão disponibilizados abertamente. Em geral, instituições investigam o perfil de usuários de dados, identificando seus interesses, bem como a utilidade social de dados e o potencial econômico que esses podem alcançar;
- **disponibilização de dados** – abrange a publicação de dados abertos de pesquisa em *sites* institucionais, plataformas de catálogo de dados abertos ou outras soluções tecnológicas. É fundamental que usuários desses sistemas consigam pesquisar conjuntos de dados, consultar metadados e fazer *downloads* de dados em formatos abertos;
- **divulgação dos dados abertos** – estratégias utilizadas para informar potenciais usuários sobre a disponibilidade de dados. Para isso, instituições podem utilizar recursos de divulgação de redes sociais e em *sites* da Internet;
- **manutenção de dados** – o esforço de gestores em prol da abertura de dados não é finalizado após disponibilização de dados e na adoção de estratégias de divulgação. Isso porque, deve haver um compromisso institucional na qual dados se manterão confiáveis, íntegros, disponíveis, atualizados e com qualidade no decorrer do tempo para que possam ser reutilizados.

Quanto à coleta de dados de pesquisa na área de Química, nota-se que o caderno de laboratório ocupa um lugar de destaque na produção científica. Isso porque tal instrumento é o objeto físico ou digital que registra os principais procedimentos executados na pesquisa, detalhando os experimentos que, por sua vez, serão posteriormente consultados e terão dados validados ou refutados, conforme sistematização minuciosa dos métodos adotados na pesquisa.

A centralidade do caderno de laboratório relacionada à comprovação de dados também foi percebida no estudo de Chen e Wu (2017), cujo objetivo foi caracterizar dados de pesquisa gerados na área de Química e descrever necessidades quanto à gestão de dados. Para isso, os autores aplicaram um questionário para 119

pesquisadores e estudantes de pós-graduação, integrantes do Instituto de Química do *Chinese Academy of Science* (CAS), no período de 2015 a 2016.

Dentre os resultados obtidos no estudo de Chen e Wu (2017), destaca-se que quase 80% dos respondentes afirmaram produzir dados experimentais. Além disso, cerca de 81% dos participantes do estudo declararam que o caderno de laboratório impresso é o instrumento mais utilizado para o registro de dados. É interessante comentar, ainda, que o caderno de laboratório em formato eletrônico e sistemas de gestão de dados aplicados a laboratórios não são canais amplamente utilizados pelos respondentes da pesquisa.

Como visto, pesquisadores na área de Química lidam rotineiramente com a produção de dados experimentais. De acordo com a Universidade de Leeds⁷⁵, geralmente, esses dados são coletados em laboratórios e, com frequência, são replicados para a verificação de padrões e de inconsistências. De maneira semelhante, Silva (2019) comentou que dados experimentais resultam de experimentos realizados em local controlado, a fim de comprovar ou rejeitar hipóteses sobre determinado fenômeno.

Outro aspecto verificado no estudo foi que pesquisadores na área da Química, geralmente, lançam mão de seus próprios padrões para descrever os conjuntos de dados que manipulam, criando assim, categorias que facilitam a recuperação desses dados.

Conforme aponta o modelo do ciclo de vida de dados criado por *United States Geological Survey* (USGS), em 2013, o detalhamento sobre dados por meio da apresentação de metadados é fundamental para que outros cientistas possam estudar, replicar e comprovar resultados de pesquisas. De modo geral, o uso de metadados padronizados, juntamente com a documentação relativa a dados, contribuem para a compreensão do contexto de produção da pesquisa científica.

Assim também, informações localizadas no *site* da *OpenAIRE*⁷⁶ demonstram que metadados são úteis para tornar conjuntos de dados facilmente localizáveis e

75 Disponível em:

https://library.leeds.ac.uk/info/14062/research_data_management/61/research_data_management_explained

76 Disponível em: <https://www.openaire.eu/what-is-metadata>

recuperáveis. Nesse contexto, metadados podem abranger informações tais como coordenadas cartográficas, descrição de instrumentos e protocolos utilizados.

Complementarmente, para a Universidade de Oulu⁷⁷, localizada na Finlândia, metadados podem ser entendidos como informações estruturadas e legíveis por máquinas sobre conjuntos de dados. Por sua vez, esses deveriam dispor de metadados e documentação suficientes para dar-lhes sentido. Logo, a principal função dos metadados é facilitar a localização de dados de pesquisa.

De maneira semelhante, nota-se que documentar procedimentos relacionados à execução de pesquisas na área de Química está associado à verificação de registros para a comprovação de dados e consequente validação do conhecimento científico. Isso porque tais registros dão margem para a checagem da fidedignidade de pesquisas, por meio da replicação de processos, e para a avaliação da qualidade de dados, conforme parâmetros sistemáticos de referência.

A associação entre dados e a possibilidade de comprovação de estudos científicos é percebida na definição de dados de pesquisa proposta por grupo multissetorial do *Research Councils UK* (RCUK), em 2016. Sob essa perspectiva, dados de pesquisa são entendidos como evidências que consolidam os resultados obtidos em um estudo, ou seja, são instrumentos de validação do conhecimento científico produzido.

De outro vértice, dados podem incluir um vasto repertório de informações presentes em depoimentos de indivíduos, descrições de experimentos, notas de observação. Também estão dispostos em diferentes formatos, incluído campos de tabelas, gravações de som e vídeo, coleções de imagens digitais, entre outros. Porém, o mais importante a dizer é que a função principal de um conjunto de dados de pesquisa é disponibilizar a informação precisa para sustentar ou validar resultados de projetos de pesquisa (RCUK, 2016)

A validação de resultados e a replicação de estudos com o objetivo de testar hipóteses é uma das características de Ciências exatas, *hard* ou nomotéticas, conforme o entendimento de Wazlawick (2014) sobre a classificação das Ciências.

77 Disponível em: https://libguides oulu.fi/Researchdata/Data_documentation

Dentre suas principais características, as ciências exatas dedicam-se ao estudo de leis previsíveis, na medida em que experimentos são reproduzidos e, geralmente, produzem os mesmos resultados, tal como ocorre na área da Física e da Química. Quanto ao rigor dos métodos utilizados, as ciências *hard* descrevem suas observações e experimentos minuciosamente, de modo a replicá-los inúmeras vezes. Por seu turno, ciências nomotéticas investigam fenômenos que se repetem, contribuindo para a descoberta de leis gerais (WAZLAWICK, 2014)

Com base em estudos como o de Borgman (2015) e o de Sayão e Sales (2015), pode-se afirmar que não há um significado específico que seja suficiente para expressar as diversas perspectivas nas quais dados de pesquisa podem ser compreendidos. Entretanto, dentre as inúmeras contribuições trazidas por esses insumos informacionais à pesquisa científica, pode-se destacar a possibilidade de comprovação de dados, a replicação de resultados obtidos e a integração de dados provenientes de outros estudos.

Nesse contexto, percebe-se que repositórios de dados podem ser vistos como portfólios da produção científica de determinada área do conhecimento, viabilizando novos canais para que dados sejam recuperados, analisados e verificados quanto ao nível de qualidade que apresentam e ao grau de confiança que inspiram.

Na área da Química, observou-se que dados são compartilhados por pesquisadores por meio de repositórios ou outras plataformas digitais, bem como por mídias sociais. Todavia, convém destacar que muitos repositórios institucionais de dados abertos de pesquisa ainda se encontram na fase de criação ou desenvolvimento, no contexto de universidades brasileiras. De certo modo, isso obstaculiza o intercâmbio de dados, uma vez que tais repositórios facilitam a inserção, recuperação, armazenamento e reuso de dados no ambiente digital, como observou Pinfield (2009).

Benefícios trazidos por repositórios de dados de pesquisa também foram apontados por Sayão e Sales (2016), dentre os quais se destacam dois aspectos: a visibilidade e o compartilhamento de dados. O primeiro ressalta que somente uma pequena parcela de dados produzidos em uma pesquisa é disponibilizada no âmbito de canais formais. Sendo assim, repositórios ampliam a visibilidade de dados, contribuindo para que esses sejam recuperados e reutilizados. Já o segundo aspecto,

refere-se à capacidade de repositórios em promover o compartilhamento de dados, agregando diferentes recursos informacionais, e, por conseguinte, fortalecendo a socialização entre membros da comunidade científica.

O presente estudo ainda observou que muitos pesquisadores de Química adotam estratégias locais para a preservação dos dados que produzem. Uma vez que dados preservados possibilitam ser consultados, e até mesmo analisados quanto à veracidade das informações que apresentam, a falta de iniciativas institucionais robustas nessa direção compromete o reuso de dados a longo prazo.

Semelhantemente, estudo de Chen e Wu (2017) no âmbito do Instituto de Química do *Chinese Academy of Science* (CAS) concluiu que a preservação de dados é realizada, prioritariamente, pelo próprio responsável da pesquisa, não havendo uma política de preservação de dados estabelecida pela instituição em que atua. Nesse sentido, a maioria dos pesquisadores e estudantes de pós-graduação (81,51% dos respondentes) utilizam computadores pessoais para preservar seus dados, 74,79% lançam mão de cadernos de laboratório impressos para esse fim e 52,10% dispõem de discos de armazenamento de dados.

Chen e Wu (2017) ressaltaram que apesar de cômodo, o armazenamento de arquivos de dados em computadores e mídias físicas não exclui o risco de perda de dados devido à obsolescência de aparelhos tecnológicos. Além disso, os autores constataram que somente 2,52% dos respondentes depositam dados em repositórios, com o objetivo de assegurarem que os mesmos sejam preservados.

O risco de dados não serem armazenados devidamente e, por conseguinte, serem perdidos ou descontinuados também foi observado no estudo de Abduldayan *et al.* (2021) que analisou práticas de gestão de dados de pesquisadores da Química em universidades de tecnologia federais localizadas na Nigéria.

Dentre os resultados alcançados na pesquisa de Abduldayan *et al.* (2021), destaca-se que pesquisadores relataram já ter perdido dados, definitivamente, em razão do armazenamento inadequado dos mesmos ou da falta de cópias de segurança. Além disso, tais pesquisadores demonstraram receio de que dados produzidos fossem acessados por indivíduos não autorizados. Para atenuá-lo, mencionaram a biblioteca como um lugar seguro para a preservação de dados a

longo prazo, bem como atribuiu-lhe a função de oferecer treinamentos sobre o tratamento de dados.

Tal como observado por Uzwyshyn (2016), repositórios de dados *online* mostram-se aliados à preservação digital, possibilitando a gestão, compartilhamento, revisão e validação de dados. Isso porque repositórios podem publicar até os chamados “dados negativos”, que são aqueles produzidos em experimentos que não obtiveram êxito na conclusão da pesquisa. Tais dados permitem que cientistas revisem determinados procedimentos e adotem caminhos metodológicos distintos, evitando assim, a duplicação de estudos.

Conforme observado neste estudo, pesquisadores da área de Química prezam por informações confiáveis e validadas por pares. Nessa direção, constatou-se que entrevistados costumam reutilizar informações oriundas de trabalhos publicados na literatura científica, e não de dados brutos, propriamente ditos. Sob essa perspectiva, o reuso de dados sem a devida atribuição autoral, geralmente, é entendido como uma ação eticamente reprovável.

Essa situação também foi percebida no estudo de Carvalho (2018) em relação ao reuso de dados de pesquisa na área da Química. Dentre os principais resultados obtidos em relação a essa prática, podem ser destacados os seguintes:

- ocorre de forma indireta, na medida em que pesquisadores utilizam documentos publicados na literatura científica;
- não é uma atividade habitual entre pesquisadores, recebendo até mesmo uma conotação negativa, em se tratando de dados ainda não publicados;
- está relacionado à replicação de resultados obtidos com o propósito de avaliar a veracidade de informações e, desse modo, distinguir quais dados são confiáveis para uso. Isso porque muitas etapas não são devidamente documentadas por pesquisadores, sendo até mesmo ocultadas para não serem revisadas posteriormente.

Mesmo com os aspectos mencionados no estudo de Carvalho (2018), nota-se que o reuso de dados oferece diversos benefícios à comunidade científica, favorecendo, inclusive, a reprodutibilidade de estudos para confirma-los ou refutá-los.

Nesse contexto, Sales *et al.* (2019) afirmam que a curadoria de dados científicos é um serviço que pode ser desempenhado por bibliotecas e que contribui para o acesso contínuo a dados e seu reuso prolongado. Por isso, é fundamental que as demandas geradas por pesquisadores em relação a dados sejam efetivamente respondidas, sendo considerados o aspecto cultural e produtivo de cada área do conhecimento.

À vista do exposto, destaca-se que este estudo observou a importância de a gestão de dados de pesquisa considerar a existência de diferenças disciplinares. De acordo com Costa (2008), as disciplinas demandam abordagens diferentes em relação ao desenvolvimento de serviços de informação, em especial, à configuração de coleções de documentos e dados dispostos em ambiente digital. Por isso, a autora enfatiza que “quem quer que tenha de planejar e operar qualquer serviço de informação, tem que levar em conta diferenças disciplinares como uma questão vital”.

Dentre os aspectos relevantes apontados por pesquisadores na área de Química, está a “propriedade intelectual em relação a dados de pesquisa”. Considerando tais dados como elementos que comprovam a realização de um estudo científico, é muito importante que haja um parâmetro expresso das condições de seu compartilhamento, uso e reprodução. Isso devido, principalmente, à produção de dados de natureza tecnológica, os quais necessitariam de condições especiais quanto aos critérios de sua disponibilização.

De acordo com Jisc⁷⁸, o reuso de dados deve ocorrer em conformidade com termos e licenças associados a dados, respeitando os direitos de responsabilidade intelectual devido a autores. Sobre as licenças de dados, Ball (2014), afirma que essas são meios utilizados para comunicar as permissões a respeito de dados a pesquisadores que, possivelmente, irão reutilizá-los. Sendo assim, uma licença pode ser entendida como um dispositivo legal, no qual uma entidade que detém direitos sobre uma obra original permite que outros a utilize, sem riscos de infração aos direitos concedidos.

Ainda segundo Ball (2014), três requisitos se destacam no contexto da concessão de licenças: 1) atribuição – reponsabilização autoral pela obra ao

⁷⁸ Disponível em: <https://rdmtoolkit.jisc.ac.uk/research-data-lifecycle/>

licenciante, quando a mesma for distribuída ou deriva nova obra; 2) *copyleft* – em trabalhos derivados do licenciado, exige-se a adoção da mesma licença; 3) não comercialidade – objetiva impedir a exploração comercial de uma obra pelo licenciado.

De acordo com Silva (2019), há diversos tipos de licenças relacionadas ao ambiente digital e a padrões de dados abertos de pesquisa, tais como *Creative Commons* e *Open Data Commons*. Na primeira são definidos os termos que condicionam o uso de obras. Já a segunda, é resultado de uma iniciativa da *Open Knowledge Foundation* (OKF) e lança mão de diferentes instrumentos jurídicos relacionados ao uso de dados abertos.

Segundo Sayão e Sales (2015), a licença *Creative Commons* é muito atribuída a conjunto de dados de pesquisa, de modo a declarar os direitos de reutilização sobre os mesmos. Todavia, essa licença não é específica a dados, havendo outras mais apropriadas para esse fim, como a licença *Open Data Commons*.

4.3 Descrever fatores que influenciam a abertura de dados de pesquisa, segundo pesquisadores em diferentes áreas

Os fatores que influenciam a abertura de dados de pesquisa na área de Sociologia e de Química serão apresentados de acordo com os resultados atingidos em cada um dos processos inerentes à teoria fundamentada, e, especificamente, nessa ordem: 1) codificação aberta, 2) codificação axial, 3) codificação seletiva e teoria substantiva. Destaca-se, ainda, que a Sociologia será a primeira área a ser abordada, seguida da Química.

4.3.1 Codificação aberta - Sociologia

Esta seção apresentará as dimensões, categorias e propriedades identificadas por meio do processo de codificação aberta, que por sua vez, foi realizado com base nos principais aspectos abordados pelos pesquisadores na área de Sociologia entrevistados no estudo. Visando anonimizar os pesquisadores entrevistados, quando esses forem mencionados no texto, serão citados pelo código “PSn^o”, em que “P” refere-se ao termo “pesquisador”, “S” ao termo “Sociologia” e “n^o” ao número sequencial atribuído a cada pesquisador, de acordo com a ordem de realização das entrevistas.

4.3.1.1 Codificação aberta da entrevista com PS1

Com relação à abertura de dados de pesquisa, PS1 afirmou que está disposto a compartilhar os dados que produz, desde que receba apoio institucional que viabilize tal prática. No que diz respeito a fatores que estimulam a abertura de dados na comunidade de pesquisadores, PS1 destacou a noção de pertencimento do pesquisador à área que atua, o qual pode ser diretamente associado ao fortalecimento do campo científico, como é possível explicitar:

O pertencimento à área, ou seja, quanto mais a comunidade acadêmica trocar materiais, a área ficará mais forte. A ideia de você ter uma comunidade científica mais robusta, que interaja a mais é o maior estímulo.

Quanto aos fatores que inibem a abertura de dados na comunidade de pesquisadores, PS1 apontou os riscos de a disponibilização de dados não respeitar aspectos éticos, além de custos jurídicos e financeiros envolvidos nesse processo.

Finalmente, PS1 relatou ser favorável à possibilidade de a universidade em que atua, agências de fomento e editores científicos tornarem obrigatório o depósito de dados abertos de pesquisa. Porém, demonstrou-se cético em relação a alguma dessas instituições viabilizarem o depósito mencionado, tal como pode-se notar:

Eu sou favorável, eu adoraria, mas quero ver quem paga a conta [...] o custo disso não é barato!

O Quadro 47 apresenta as dimensões, categorias e propriedades identificadas por meio de codificação aberta na entrevista com PS1.

Quadro 47 – Codificação aberta da entrevista com PS1: fatores que influenciam a abertura de dados de pesquisa

Dimensão	Categoria	Propriedade
Disposição para compartilhar dados	Favorável	Compartilhamento aberto de dados de pesquisas
Fator estimulador à abertura de dados	Pertencimento à área	Colaboração entre a comunidade acadêmica
Fator inibidor à abertura de dados	Riscos éticos	Infrações éticas
	Custos	Responsabilidade sobre custos
Obrigatoriedade pela universidade	Favorável	Possibilidade de tornar obrigatório
Obrigatoriedade por agências de fomento	Favorável	Possibilidade de tornar obrigatório
Obrigatoriedade por editores científicos	Favorável	Possibilidade de tornar obrigatório

Fonte: elaborado pela autora

4.3.1.2 Codificação aberta da entrevista com PS2

No que se refere à abertura de dados de pesquisa, PS2 disse que está disposto a compartilhar dados de pesquisa, mas reconheceu que sendo uma prática inerente ao contexto da Ciência aberta, esse tema necessita ser profundamente debatido por diversas instituições, como pode-se notar:

Acho que essa questão de Ciência aberta requer um debate, debate profundo feito pelas intuições, pela participação das universidades públicas, que geram 80% das pesquisas neste país. Mas também tem que dar espaço para os institutos, para agências, para representações, para os conselhos de reitores.

Ao falar sobre fatores que estimulam a abertura de dados de pesquisa, PS2 destacou a possibilidade de aprimoramento da própria pesquisa, que passa ter outros dados ao seu dispor, até mesmo mais atualizados, que podem contribuir para retratar aspectos da sociedade. Quanto ao que poderia inibir tal abertura, PS2 apontou a apropriação indevida de dados, na qual alguns pesquisadores poderiam suprimir a identidade do verdadeiro produtor do dado, visando obter vantagens no âmbito científico.

Com relação à possibilidade de a universidade em que atua tornar obrigatório o depósito de dados abertos de pesquisa, PS2 disse que ainda não há garantias de que a gestão de dados será adequada, uma vez que não há controle institucional sobre esses. Semelhantemente, ele apontou que não há discussões ou diretrizes representativas sobre tal obrigatoriedade no âmbito de agências de fomento. Já em relação aos editores científicos, PS2 ressaltou que o depósito de dados é fundamental para atestar a veracidade de fontes que uma vez foram contestadas. Nesse sentido, explicou:

Eu acho que se o editor científico contesta os teus dados, porque acha que os teus dados não estão corretos, tu tens a obrigação de mostrar que os dados foram levantados. Pode questionar a tua amostragem que não foi representativa, o que é uma informação isolada que não reflete pensamento dentro da distribuição normal da curva da causa e por aí fora.

O Quadro 48 é resultado do processo de codificação aberta e apresenta as dimensões, categorias e propriedades identificadas no relato de PS2.

Quadro 48 – Codificação aberta da entrevista com PS2: fatores que influenciam a abertura de dados de pesquisa

Dimensão	Categoria	Propriedade
Disposição para compartilhar dados	Debates	Discussão
Fator estimulador à abertura de dados	Aprimoramento da pesquisa	Retrato consistente da sociedade
Fator inibidor à abertura de dados	Apropriação indevida de dados	Omissão de autoria para o recebimento de vantagens
Obrigatoriedade pela universidade	Garantias sobre a gestão de dados	Falta de controle sobre dados
Obrigatoriedade por agências de fomento	Diretrizes	Falta de orientações para o depósito de dados de pesquisa
Obrigatoriedade por editores científicos	Comprovação de dados	Verificação da fidedignidade de dados

Fonte: elaborado pela autora

4.3.1.3 Codificação aberta da entrevista com PS3

Ao falar sobre abertura de dados de pesquisa, PS3 demonstrou bastante disposição em compartilhar dados que produz. Dentre os fatores que poderiam estimular a abertura de dados em sua área de atuação, foi destacada a obrigatoriedade estabelecida por agências de fomento para que pesquisadores disponibilizem abertamente dados coletados a partir de recursos concedidos. Já em relação a fatores inibidores a essa prática, PS3 ressaltou o receio de pesquisadores perderem o ineditismo da publicação de dados que geraram em suas pesquisas. Assim, PS3 comentou:

Eu acho que o que inibe é essa percepção de que se você ceder esses dados, você vai perder a possibilidade de você falar sobre eles, ou você não vai ser a única pessoa falando sobre eles. Então, tem essa enorme preocupação, na minha opinião, de ser inédito, né? E de ninguém mais falar sobre esse assunto.

PS3 mostrou-se favorável à possibilidade de a universidade em que atua tornar obrigatório o depósito de dados abertos de pesquisa. Isso, com base no argumento de que, por se tratar de estudos produzidos com financiamento público, os dados gerados deveriam retornar à sociedade de algum modo. Ao se referir às agências de fomento, PS3 salientou a falta de iniciativas que requeiram a abertura de dados de pesquisa, o que sugere a necessidade de uma liderança institucional mais sensível a esse contexto. Nessa direção, PS3 explicou:

Na verdade, hoje em dia, as agências de fomento sequer pedem os nossos bancos de dados. Quando a gente ganha um financiamento público e depois a gente vai prestar contas, o que eles pedem é um relatório, e depois, que você indique as produções no seu currículo e no currículo da equipe associada àquele projeto. E é isso. Então, não há nenhum pedido, inclusive do ponto de vista da preservação, né? [...] eu acho que seria, inclusive, interessante que as agências de fomento tivessem um papel maior de liderança, ao invés de obrigar o pesquisador, por exemplo, de disponibilizar os dados, que ela própria disponibilizasse, que ela pedisse toda essa documentação, mas que depois, no final do ano, criasse um site. Olha! Está aqui todas as bases de dados e todos os arquivos produzidos pelos pesquisadores!

Com relação à possibilidade de editores científicos tornarem obrigatório o depósito de dados abertos de pesquisa, PS3 disse que, cada vez mais, tais editores têm solicitado conjuntos de dados, e que essa medida é interessante do ponto de vista da replicabilidade da ciência. Logo, possui um caráter comprobatório, diretamente relacionado à verificação da fidedignidade de dados que serão publicados.

O Quadro 49 é resultado do processo de codificação aberta e mostra as dimensões, categorias e propriedades assinaladas a partir da entrevista com PS3.

Quadro 49 – Codificação aberta da entrevista com PS3: fatores que influenciam a abertura de dados de pesquisa

Dimensão	Categoria	Propriedade
Disposição para compartilhar dados	Favorável	Compartilhamento de dados abertos de pesquisas
Fator estimulador à abertura de dados	Obrigatoriedade	Agência de financiamento à pesquisa
Fator inibidor à abertura de dados	Perda de ineditismo	Publicação antecipada
Obrigatoriedade pela universidade	Favorável	Possibilidade da obrigatoriedade
Obrigatoriedade por agências de fomento	Falta de iniciativa	Falta de liderança
Obrigatoriedade por editores científicos	Comprovação de dados	Verificação da fidedignidade de dados

Fonte: elaborado pela autora

4.3.1.4 Codificação aberta da entrevista com PS4

PS4 declarou que tem disposição para compartilhar dados de pesquisas, contanto que haja recursos suficientes e efetivos para viabilizar tal prática. Quando comentou sobre fatores que poderiam estimular a abertura de dados na área em que atua, PS4 salientou a conscientização de pesquisadores sobre a importância dessa abertura, no que se refere à colaboração científica.

Quanto aos fatores inibidores à abertura de dados de pesquisa, PS4 destacou a falta de recursos, o que implica diretamente na disponibilização de infraestrutura adequada para viabilizar essa prática. A partir de sua fala, nota-se que pesquisadores brasileiros se encontram sobrecarregados com as rotinas acadêmicas e sem apoio de infraestrutura nas instituições nas quais atuam.

Você tem que cuidar dos cursos de graduação, não é só dar as suas aulas. Você tem que fazer a administração de tudo isso, e, simplesmente, não tem infraestrutura nenhuma para isso. Nenhuma! Nenhuma! Você não tem secretarias, não tem gestão de projetos, não tem gestão financeira, não tem nada. Então é um desastre do ponto de vista de infraestrutura no Brasil.

Mesmo declarando-se favorável à possibilidade de a universidade em que atua tornar obrigatório o depósito de dados abertos de pesquisa, PS4 evidenciou ser necessário dispor de condições estruturais para isso. Ademais, sugeriu que

universidades criassem mecanismos de recompensas a pesquisadores adeptos ao compartilhamento de dados abertos. Como apontou, poderia haver um sistema de pontuação que gratificasse pesquisadores que mais disponibilizassem dados, concedendo-lhes aumento do recurso destinado ao financiamento de pesquisas:

Agora, seria muito diferente se dissessem: 'Olha! A partir desse momento em que você disponibilizou todos os seus dados, você vai ganhar acesso a esse tipo de projeto de pesquisa, que é um acesso aos grandes projetos de pesquisas'. Você está aqui, mas você não disponibiliza os seus dados, então, você perde ponto para chegar nos financiamentos maiores, ou se você disponibilizar os seus dados, você vai ganhar 20% a mais no próximo projeto de pesquisa que você for aprovado, para poder estruturar a disponibilização de dados.

No caso da possível obrigatoriedade do depósito de dados de pesquisa instituída por agências de fomento, PS4 não viu muitos entraves, uma vez que considera que tais agências detêm o dinheiro que pode viabilizar os recursos necessários para essa prática. Com relação ao contexto de editores científicos, PS4 salientou que o foco desses ainda está em publicações tradicionais, como artigos de periódicos, e não em dados de pesquisa.

O Quadro 50 é resultado do processo de codificação aberta e apresenta as dimensões, categorias e propriedades identificadas a partir da entrevista com PS4.

Quadro 50 – Codificação aberta da entrevista com PS4: fatores que influenciam a abertura de dados de pesquisa

Dimensão	Categoria	Propriedade
Disposição para compartilhar dados	Recursos	Disponibilidade de recursos
Fator estimulador à abertura de dados	Conscientização de pesquisadores	Importância da abertura de dados de pesquisa
Fator inibidor à abertura de dados	Falta de infraestrutura	Falta de recursos
Obrigatoriedade pela universidade	Condições estruturais	Disponibilidade de recursos
	Falta de incentivos	Falta de estratégias de recompensa
Obrigatoriedade por agências de fomento	Disponibilidade de recursos	Estrutura para a abertura de dados de pesquisa
Obrigatoriedade por editores científicos	Publicação de artigos	Foco na publicação

Fonte: elaborado pela autora

4.3.1.5 Codificação aberta da entrevista com PS5

PS5 afirmou que tem disposição para compartilhar dados de pesquisas, entretanto, nem todos, em razão de alguns tratarem de temáticas sensíveis que eticamente não poderiam ser expostas. Sobre esse aspecto, também explicou que a disponibilização de dados abertos dependia da instituição e da natureza da pesquisa:

Depende da instituição e depende também da natureza da pesquisa. Eu acho que tem entrevistas que eu tenho que eu não disponibilizaria porque tem nomes de ministros, tem nomes de governantes, de políticos, você entende? Assim, são dados bem sensíveis, então ficar aberto fica complicado, porque as pessoas confiaram que eu ia usar aquela entrevista para minha pesquisa sem divulgar o nome delas, sem identificá-las. Eu iria identificar o processo que elas estão envolvidas, processo, assim: de formação de mentalidade; de escolhas; de pessoas; enfim! Então, eu acho que é uma questão que tem que ver mais de perto, certo?

Com relação aos fatores estimuladores à abertura de dados de pesquisa, PS5 salientou a importância da ascensão de lideranças institucionais que orientem como a gestão de dados de pesquisa deve ser realizada. Quanto aos fatores inibidores, PS5 ressaltou que a abertura de dados deve ser algo facultativo ao pesquisador, e não obrigatório. Portanto, a disponibilização obrigatória de dados por parte do pesquisador não foi indicada como uma medida eficaz para promover a abertura de dados.

Ao ser questionado sobre a possível obrigatoriedade do depósito de dados de pesquisa instituída por universidades, agências de fomento e editores científicos, PS5 declarou-se contrário a tal medida, por entender que o pesquisador tem livre-arbítrio para decidir se deseja, ou não, disponibilizar dados abertos. Referindo-se ao contexto de universidades, PS5 teceu o seguinte comentário:

Que não seja de forma alguma por constrangimento. Eu acho que nenhum pesquisador tem a obrigação de disponibilizar o dado dele para outras pesquisas. Ele pode ter planos de longo prazo com algum tipo de trabalho científico, né? Mas à medida que ele achar que pode, fica à vontade a universidade fazer isso.

O Quadro 51 é resultado do processo de codificação aberta e mostra as dimensões, categorias e propriedades assinaladas com base no relato de PS5.

Quadro 51 – Codificação aberta da entrevista com PS5: fatores que influenciam a abertura de dados de pesquisa

Dimensão	Categoria	Propriedade
Disposição para compartilhar dados	Dados sensíveis	Ética da pesquisa
Fator estimulador à abertura de dados	Liderança institucional	Gestão de dados de pesquisa institucionalizada
Fator inibidor à abertura de dados	Obrigatoriedade	Disponibilização obrigatória de dados
Obrigatoriedade pela universidade	Obrigatoriedade	Livre consentimento do pesquisador
Obrigatoriedade por agências de fomento		
Obrigatoriedade por editores científicos		

Fonte: elaborado pela autora

4.3.1.6 Codificação aberta da entrevista com PS6

PS6 afirmou que tem disposição para compartilhar dados de pesquisa qualitativos ou quantitativos, contanto que não sejam dados sensíveis, tais como os que poderiam colocar a vida de pessoas em risco. Também destacou que o aprimoramento de pesquisa é um fator que pode estimular a abertura de dados de pesquisa, em especial, a possibilidade de fazer análises comparadas em regiões geográficas distintas. No que concerne a fatores inibidores, PS6 apontou a exposição a riscos éticos, quando se trata de dados sensíveis. Sistemáticamente, tais dados demandam anonimização de informações sobre indivíduos, eventos e locais.

Ao opinar sobre a possibilidade de a universidade em que atua e agências de fomento tornarem obrigatório o depósito de dados de pesquisa, PS6 decidiu-se por uma via dialógica, na qual pesquisadores possam adentrar no processo de conhecimento dos princípios epistemológicos da Ciência aberta e optar pela disponibilidade de dados, ou não:

É... obrigatório eu já não gosto. Eu já acho complicado porque, se seguir a tradição nossa brasileira de que não discute nada e depois torna um projeto obrigatório, a resistência a ele vai ser gigantesca. Então, eu preferiria um caminho de construção, igual a gente tem consolidado no Brasil o currículo lattes, que a gente coloca as informações ali e tal. A gente teria que ter esse processo de conhecimento das bases epistemológicas da Ciência aberta, para que nós, enquanto cientistas, estivéssemos convencidos nisso. Então eu optaria pela linha do convencimento nesse sentido para a universidade, assim como as agências.

Já no contexto de editores científicos, PS6 descreveu um cenário de incertezas quanto à disponibilização de dados. Nesse sentido, não se sabe se os mesmos serão utilizados para corroborar a publicação de artigos ou disponibilizados separadamente. O Quadro 52 aborda as dimensões, categorias e propriedades identificadas na codificação aberta, conforme o relato de PS6.

Quadro 52 – Codificação aberta da entrevista com PS6: fatores que influenciam a abertura de dados de pesquisa

Dimensão	Categoria	Propriedade
Disposição para compartilhar dados	Dados sensíveis	Aspectos éticos
Fator estimulador à abertura de dados	Aprimoramento da pesquisa	Possibilidade de fazer análises comparadas em regiões geográficas distintas
Fator inibidor à abertura de dados	Riscos éticos	Disponibilização indevida de dados de pesquisa
Obrigatoriedade pela universidade	Livre escolha do pesquisador	Depósito não obrigatório
Obrigatoriedade por agências de fomento		
Obrigatoriedade por editores científicos	Cenário de incertezas	Indefinições quanto à disponibilização de dados

Fonte: elaborado pela autora

4.3.1.7 Codificação aberta da entrevista com PS7

PS7 demonstrou-se favorável à disponibilização de dados abertos que produz, mas para isso, seria imprescindível uma abertura institucionalizada, na qual a universidade em que atua pudesse assegurar a não exposição de dados a riscos de privacidade. Sobre isso, PS7 explicou:

Eu acho que se fosse uma pesquisa quantitativa não teria problema. O receio é dentro de um quadro em que a universidade da qual eu faço parte não tem uma política desse tipo. Abrir a minha pesquisa individualmente é me expor a um risco, e eu não sei se eu precisaria ter uma garantia institucional da universidade para que isso pudesse ser feito. Então, eu apesar de achar importante fazer isso, é preciso que isso seja institucionalizado antes da abertura.

Para PS7, o aprimoramento de pesquisas por meio da análise de novos dados é um fator que pode estimular a abertura de dados de pesquisa, enquanto a não institucionalização de tal abertura apresenta-se como um fator inibidor. Além disso, ao comentar a respeito da possibilidade de a universidade em que atua, agências de

fomento e editores científicos tornarem obrigatório o depósito de dados de pesquisa, PS7 declarou-se contrário a qualquer arbitrariedade nesse sentido.

Porém, ressaltou que caso existisse uma hierarquia, acreditava que as revistas seriam as primeiras a adotarem a obrigatoriedade, devido a existência de políticas próprias em cada comitê editorial. Em seguida, viriam as agências de fomento, que talvez estabeleceriam o depósito obrigatório como requisito para a aprovação de projetos. Por fim, viriam as universidades, sobre as quais fez as seguintes declarações:

Sou completamente contrário. Eu acho que isso é uma arbitrariedade, seria uma arbitrariedade da própria universidade, e como eu disse, se for uma coisa acordada coletivamente dentro da comunidade acadêmica e etc., etc., faz sentido. Senão, se for simplesmente: 'a partir de 31 de dezembro tem que abrir os dados, eu acho que não'. Eu acho que teria que ser uma coisa democraticamente construída. [...] o professor que não quiser abrir os dados, ele não abre, tem que ser uma coisa assim: ele pode aderir ou não, faz sentido para ele isso? Faz sentido para ele? Ele não quer abrir os dados porque ele acha que os dados ainda são parciais, se ele abrir os dados alguém pode utilizar de maneira equivocada? Bom, tem zilhões de consequências que a gente pode pensar sobre abertura dos dados.

O Quadro 53 apresenta as dimensões, categorias e propriedades identificadas na codificação aberta, com base na entrevista com PS7.

Quadro 53 – Codificação aberta da entrevista com PS7: fatores que influenciam a abertura de dados de pesquisa

Dimensão	Categoria	Propriedade
Disposição para compartilhar dados	Institucionalização	Abertura de dados institucionalizada
Fator estimulador à abertura de dados	Aprimoramento da pesquisa	Desenvolvimento científico
Fator inibidor à abertura de dados	Institucionalização	Abertura não institucionalizada de dados
Obrigatoriedade pela universidade		
Obrigatoriedade por agências de fomento	Obrigatoriedade	Arbitrariedade
Obrigatoriedade por editores científicos		

Fonte: elaborado pela autora

4.3.1.8 Codificação aberta da entrevista com PS8

PS8 afirmou estar disposto a compartilhar dados das pesquisas que realiza, entretanto, destacou ser importante que produtores de dados tenham a primazia de disponibilizá-los antes de outros pesquisadores. Nesse sentido, PS8 recomendou a definição de um período para que os responsáveis de determinada pesquisa possam

ter exclusividade inicial para trabalhar com os dados gerados, e apenas após a vigência desse prazo, tais dados estariam disponíveis abertamente.

Ao abordar fatores que estimulariam a abertura de dados entre pesquisadores, PS8 destacou a obrigatoriedade do depósito de dados que foram gerados mediante financiamento público, após determinado prazo. Com relação a aspectos inibidores à tal abertura, PS8 foi enfático ao mencionar a vaidade de pesquisadores e o apego desses no que refere aos dados que produzem, como se pode observar:

O que pode inibir? É a vaidade. Por exemplo, algo muito subjetivo: 'eu fiz, é meu, e eu quem falei que vou mexer, quem eu quiser que mexa nesses dados, mais ninguém'. Então, tem um controle muito forte em cima disso, e, muitas vezes, alguns pesquisadores fazem isso, mesmo tendo sido dinheiro público. Que aí eu acho que você tem um ponto complicado, você tem o seu trabalho, o seu esforço intelectual, mas, por outro lado, isso foi possibilitado com dinheiro público. Como chegar o meio termo, aí?

Outro aspecto inibidor apontado por PS8 foi a publicação de dados abertos e o seu acesso mediante pagamento. De maneira ilustrativa, PS8 comentou que, muitas vezes, o custo para publicar um artigo em acesso aberto é muito alto, chegando a doze mil dólares ou mais. Em sua perspectiva, isso praticamente inviabiliza a publicação científica brasileira no contexto internacional, em razão da insuficiência de recursos destinadas a esse fim.

Ao comentar sobre a possibilidade de a universidade em que atua tornar obrigatório o depósito de dados, PS8 comentou a respeito da ausência de normativos, ou uma legislação que possa vigorar como mecanismo legal que contenha tal obrigatoriedade. Assim, as universidades, por si, não poderiam estabelecer essa cobrança a pesquisadores, sem antes investigarem a constitucionalidade dessa medida.

Quanto à obrigatoriedade no contexto de agências de fomento, PS8 disse que consultando a legalidade de tal medida, pesquisas realizadas mediante financiamento público poderiam requerer isso a pesquisadores. Todavia, seria necessário firmar acordos de disponibilização de dados, em que se oficializaria sob quais termos esses dados poderiam ser liberados.

No que se refere à obrigatoriedade no contexto de editores científicos, PS8 relatou que alguns periódicos brasileiros já adotam essa medida, sendo uma prática cada vez mais realizada no contexto internacional. Em sua perspectiva, é aceitável que editores solicitem dados de pesquisadores para fim de verificação da

fidedignidade de processos de pesquisa, principalmente, quando há alguma suspeita de resultados não consolidados.

O Quadro 54 mostra as dimensões, categorias e propriedades identificadas na codificação aberta, com base na entrevista com PS8.

Quadro 54 – Codificação aberta da entrevista com PS8: fatores que influenciam a abertura de dados de pesquisa

Dimensão	Categoria	Propriedade
Disposição para compartilhar dados	Exclusividade	Período exclusivo para trabalhar com dados
Fator estimulador à abertura de dados	Financiamento público da pesquisa	Dados obtidos mediante recursos públicos
Fator inibidor à abertura de dados	Vaidade de pesquisadores	Apego a dados
	Pagamento	Publicação de dados em periódicos científicos mediante pagamento
Obrigatoriedade pela universidade	Legislação	Ausência de normas
Obrigatoriedade por agências de fomento	Financiamento público da pesquisa	Acordos/termos de disponibilização de dados
Obrigatoriedade por editores científicos	Comprovação de dados	Verificação da fidedignidade de dados

Fonte: elaborado pela autora

4.3.1.9 Codificação aberta da entrevista com PS9

PS9 afirmou estar disposto a compartilhar dados que produz. Entretanto, explicitou sua preocupação em relação à proteção dos mesmos, principalmente, quanto à anonimização de participantes de pesquisas. Nesse sentido, PS9 considerou que o tratamento de dados qualitativos é mais desafiador do que o de dados quantitativos:

Sou sim, a favor. Eu acho que isso é fundamental para a ciência, mas reconheço que teria dificuldades, principalmente, em relação à proteção de identidade de dados qualitativos. Para dados quantitativos, eu não teria nenhum problema. Geralmente, eles não trazem identificação e é fácil de você excluir uma variável que tenha [...] mas o dado qualitativo é mais difícil de anonimizar.

Ao abordar fatores que poderiam estimular a abertura de dados entre pesquisadores, PS9 mencionou que o acesso aberto é potencializado por tratativas políticas, não configurando ser uma temática tratada apenas no âmbito acadêmico. Com relação a aspectos inibidores, PS9 destacou que pesquisadores não dispõem de

habilidades técnicas suficientes para lidar com os processos relacionados à abertura de dados.

Quando opinou sobre a possibilidade de a universidade em que atua tornar obrigatório o depósito de dados abertos de pesquisa, PS9 comentou que tal medida abre portas para uma maior interação entre universidade e comunidade, uma vez que se trata de uma temática pouquíssimo discutida entre essas duas entidades. Nessa perspectiva, PS9 explicou que considera a extensão universitária uma via para a consolidação da Ciência aberta e institucionalização da abertura de dados:

Eu acho que é uma porta para isso porque aí é uma abertura da universidade para a comunidade, e é claro que o acesso aberto beneficia muito mais a comunidade, faz a universidade sair um pouco daquela representação que o pessoal faz dela, da gente, por exemplo, que nós estamos na universidade, que é um setor de filtro, que não está aberta para toda a população, como nós da torre de marfim. Então, eu acho que seria a curricularização da extensão, é uma investigativa que eu acho que ajudaria, que vai no mesmo sentido, do que a ciência aberta e essa tentativa de institucionalizar a abertura de dados de pesquisas, de compartilhamento e abertura.

Quanto à possibilidade de agências de fomento tornarem obrigatório o depósito de dados abertos de pesquisa, PS9 demonstrou ceticismo no que se refere à adesão de pesquisadores a tal requisição. Isso, em razão de algumas áreas do conhecimento poderem apresentar mais restrições à abertura do que outras. Sendo assim, PS9 comentou:

Aí eu já sou um pouco mais cético, porque tem outras áreas, por exemplo, algumas áreas que são voltadas mais à produção industrial, ou ao comércio. Então, eu tenho a impressão que pode haver uma que tenha uma restrição maior a esse tipo de abertura. Então, por parte das agências de fomento, ainda mais atualmente, nesse contexto que está atualmente o poder, eu sou meio cético.

Já no contexto de possível obrigatoriedade estabelecida por editores científicos, PS9 observa uma tendência, haja vista que a publicação de artigos já carrega consigo alguns conceitos relacionados à Ciência aberta, como a disponibilização de texto completo por meio de plataformas digitais. O Quadro 55 apresenta as dimensões, categorias e propriedades identificadas na codificação aberta, com base na entrevista com PS9.

Quadro 55 – Codificação aberta da entrevista com PS9: fatores que influenciam a abertura de dados de pesquisa

Dimensão	Categoria	Propriedade
Disposição para compartilhar dados	Proteção de dados	Anonimização de participantes
Fator estimulador à abertura de dados	Aspecto político	Acesso aberto em agenda política
Fator inibidor à abertura de dados	Habilidades técnicas	Expertise para o tratamento de dados
Obrigatoriedade pela universidade	Tema pouco debatido	Temática pouquíssimo discutida entre universidade e sociedade
Obrigatoriedade por agências de fomento	Diferenças disciplinares	Níveis de restrições à abertura de dados
Obrigatoriedade por editores científicos	Publicação de artigos	Tendência

Fonte: elaborado pela autora

4.3.1.10 Codificação aberta da entrevista com PS10

PS10 informou estar disposto a compartilhar dados de pesquisa, contanto que sejam considerados os compromissos éticos firmados com participantes da pesquisa. Dentre os fatores que poderiam estimular a abertura de dados de pesquisa, PS10 destacou a intensificação de debates sobre o fenômeno da Ciência aberta, principalmente em programas de pós-graduação e institutos de pesquisa. Isso, por considerar que a abertura de dados é um tema ainda pouco explorado na área em que atua, o que se constitui um fator inibidor nesse sentido:

Não é um tema recorrente no meu programa de pós-graduação. Por exemplo, nós nunca debatemos esse assunto, e é um programa que tem um perfil interdisciplinar, que tem uma abertura para discussão com pesquisadores de áreas distintas daquelas aonde você tem sua especialização [...] então eu acho que começaria por aí, né?

Ao opinar sobre a possibilidade de a universidade em que atua e editores científicos tornarem obrigatório o depósito dos dados das pesquisas que conduz, PS10 ressaltou que por esse ser um tópico pouco debatido, não via muitas possibilidades de implementação a curto prazo. Já com relação ao contexto de agências de fomento à pesquisa, PS10 destacou que se tal obrigação ocorresse, pesquisadores (principalmente coordenadores de projetos) ficariam sobrecarregados, como é exposto em sua fala:

Eu não levaria por surpresa porque uma hora ou outra isso apareça, surja como uma imposição. O que eu gostaria de lamentar, porque seria mais uma incumbência, exigência para coordenadores de projeto,

que já não são poucas. Então, eu reafirmo aquilo que eu disse na questão anterior, é possível que aconteça, mas isso significa mais sobrecarga de trabalho para principalmente coordenadores de projetos de pesquisa.

O Quadro 56 apresenta as dimensões, categorias e propriedades identificadas na codificação aberta, a partir da entrevista com PS10.

Quadro 56 – Codificação aberta da entrevista com PS10: fatores que influenciam a abertura de dados de pesquisa

Dimensão	Categoria	Propriedade
Disposição para compartilhar dados	Favorável	Aspectos éticos
Fator estimulador à abertura de dados	Debates	Mais discussões sobre o tema
Fator inibidor à abertura de dados	Tema pouco explorado	Programas de pós-graduação
Obrigatoriedade pela universidade	Tema pouco debatido	Pouco provável de ocorrer
Obrigatoriedade por editores científicos		
Obrigatoriedade por agências de fomento	Sobrecarga de trabalho	Sobrecarga a coordenadores de projetos

Fonte: elaborado pela autora

4.3.1.11 Síntese da codificação aberta

O processo de codificação aberta nas entrevistas realizadas resultou em uma primeira descrição dos fatores que influenciam a abertura de dados de pesquisa na área de Sociologia. Dessa forma, foram identificadas 49 categorias abertas, dispostas em seis dimensões. Tais categorias passaram por processo de codificação axial, convertendo-se em outras quatro categorias axiais, que, por seu turno, se desdobraram em dez subcategorias, as quais podem ser observadas no Quadro 57.

Quadro 57 – Síntese da codificação aberta – Sociologia: fatores que influenciam a abertura de dados de pesquisa

Dimensão	Categoria aberta	Categoria axial	Subcategoria axial
Disposição para compartilhar dados	Favorável	Fatores estimuladores à abertura de dados	Aprimoramento de estudos na área
	Debates		Financiamento público da pesquisa
	Recursos		Liderança institucional
	Dados sensíveis	Fatores inibidores à abertura de dados	Riscos éticos
			Falta de infraestrutura
			Tópico pouco explorado
			Publicação em acesso aberto mediante pagamento
	Fator estimulador à abertura de dados	Pertencimento à área	Favorável à obrigatoriedade
Aprimoramento da pesquisa		Pesquisas fomentadas com dinheiro público	
Obrigatoriedade		Desfavorável à obrigatoriedade	Livre arbítrio do pesquisador
Conscientização de pesquisadores			
Liderança institucional			
Financiamento público da pesquisa			
Aspecto político			
Debates			
Fator inibidor à abertura de dados	Riscos éticos		
	Custos		
	Apropriação indevida de dados		
	Perda de ineditismo		
	Falta de infraestrutura		
	Obrigatoriedade		
	Institucionalização		
	Vaidade de pesquisadores		
	Pagamento		
	Habilidades técnicas		
	Tema pouco explorado		
Obrigatoriedade pela universidade	Favorável		
	Garantias sobre a gestão de dados		
	Condições estruturais		
	Falta de incentivos		
	Obrigatoriedade		
	Livre escolha do pesquisador		
	Legislação		
	Tema pouco debatido		

Obrigatoriedade por agências de fomento	Favorável	
	Diretrizes	
	Falta de iniciativa	
	Disponibilidade de recursos	
	Obrigatoriedade	
	Livre escolha do pesquisador	
	Financiamento público da pesquisa	
	Diferenças disciplinares	
	Sobrecarga de trabalho	
Obrigatoriedade por editores científicos	Favorável	
	Comprovação de dados	
	Publicação de artigos	
	Obrigatoriedade	
	Cenário de incertezas	
	Tema pouco debatido	

Fonte: elaborado pela autora

4.3.2 Codificação axial – Sociologia

Esta seção relacionará as categorias e subcategorias identificadas no estudo, com base no processo de codificação axial realizado.

4.3.2.1 Categoria “Fatores estimuladores à abertura de dados de pesquisa”

A categoria “Fatores estimuladores à abertura de dados de pesquisa” representa os aspectos identificados no estudo que contribuem para incentivar a abertura de dados de pesquisa. Sob outro ângulo, trata-se do conjunto de forças que agem sobre pesquisadores, encorajando-os, de modo a influenciar a sua atitude rumo à abertura de dados.

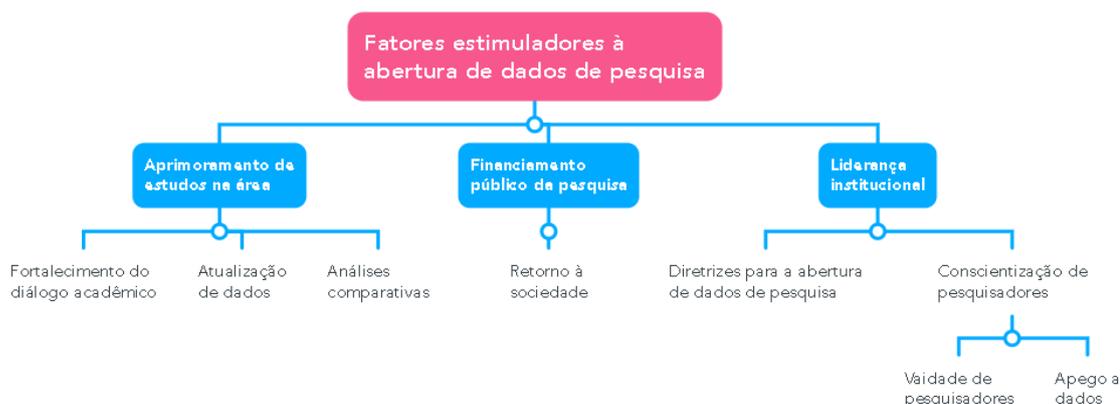
Conforme pode ser visto na Figura 41, esta categoria desenvolveu-se nos termos de três subcategorias, sendo elas: aprimoramento de estudos na área; financiamento público da pesquisa; liderança institucional. A primeira, refere-se à possibilidade de o pesquisador cooperar para o desenvolvimento da Sociologia, por meio do fortalecimento do diálogo acadêmico, a atualização de dados de estudos realizados e a realização de análises entre diferentes estudos.

A segunda subcategoria mencionada apresentou o retorno de publicações e dados à sociedade como principal argumento para a abertura de dados de pesquisa. Sob esse ponto de vista, a sociedade deveria se beneficiar da produção científica, uma vez que parcela considerável das pesquisas mencionadas no estudo foram financiadas por meio de recursos provenientes de instituições públicas.

Por fim, a terceira subcategoria referida diz respeito à ascensão de uma instituição que lidere a gestão de dados abertos de pesquisa no contexto brasileiro, de modo que sejam estabelecidas diretrizes que orientem pesquisadores sobre como proceder em relação aos dados produzidos. Ademais, tal instituição poderia adotar linhas de convencimento para conscientizar pesquisadores a respeito da importância da disponibilização de dados abertos no âmbito da comunicação científica. Isso porque, segundo entrevistados no estudo, a vaidade e o apego excessivo de alguns

pesquisadores pelos dados que produziram são aspectos que podem dificultar a abertura de dados.

Figura 41 – Categoria Fatores estimuladores à abertura de dados de pesquisa - Sociologia



Fonte: elaborado pela autora

- Subcategoria “Aprimoramento de estudos na área”

Com base nas informações obtidas nesta pesquisa, notou-se que o aprimoramento de estudos na área de Sociologia foi apontado como um dos fatores que estimulam a abertura de dados de pesquisa. Nessa direção, PS1 alegou que tal abertura contribui para o fortalecimento da área, tornando-a mais robusta teoricamente. Logo, aumenta-se a capacidade analítica sobre dados, uma vez que o produtor dos mesmos, sozinho, não conseguiria explorá-los em sua totalidade, e o conhecimento não se encerraria nesse momento. Sobre isso, PS1 comentou:

E o que mais incentiva é fortalecer a área. Fortalecer o diálogo acadêmico, ter uma comunidade acadêmica mais robusta, e junto com isso, você aumentar a capacidade analítica, porque quem produz dados em quantidade, quem é um produtor de dado, em geral, ele não dá conta de analisar tudo aquilo.

De acordo com PS2, a abertura de dados de pesquisa proporcionaria a obtenção de dados que, cada vez mais, retratariam a sociedade por trazerem descrições mais recentes sobre fenômenos observados. A título de exemplo, PS2 mencionou o tópico da fome vivenciada atualmente. Um estudo com dados atualizados sobre esse assunto enriqueceria o campo de estudo sociológico, além de publicizar tal situação para que venha a conhecimento público e, dessa forma, ser fonte orientadora nos processos de tomada de decisão. A esse respeito, PS2 disse:

Eu acho que os aspectos que animariam seria a possibilidade sempre de trazer um retrato da sociedade. No meu caso, atualizado, consistente, rigoroso que realmente retratasse a realidade, olha a questão da fome hoje, da desigualdade, então, tu vais lá e traz, como é que as pessoas estão ingerindo? Como é que elas tão fazendo para sobreviver em um cenário tão adverso? Trazer essa informação de campo é poder colocar isso e publicar.

No que se refere ao aprimoramento de estudos na área, PS6 informou que a abertura de dados de pesquisa possibilitaria a realização de análises comparativas sobre dados coletados em regiões geográficas distintas. Consequentemente, isso traria economicidade à produção de pesquisas e potencializaria outros trabalhos com a redução de tempo para a obtenção de dados. Nessa direção, PS6 comentou:

Olha, os aspectos que facilitariam, que estimulariam seriam essa ideia, no meu ponto de vista, de eu conseguir fazer análises comparadas, que é o que mais gosto de fazer. Então, se eu tenho colegas que pesquisam assuntos correlatos em outras localidades e eles disponibilizam também, isso potencializa melhor o trabalho dele, porque não vai ter que vir para cá para levantar determinadas informações que já vão estar disponíveis para ele e eu não terei que ir para lá, vamos supor, né? Ou qualquer outra coisa do tipo. Então, eu acho que isso facilitaria.

De maneira semelhante, PS3 afirmou que a abertura de dados de pesquisa é interessante para a compreensão de estudos que estão sendo realizados na área, uma vez que o processo da pesquisa científica gira em torno da complementaridade de estudos, conforme explicou:

Sim, estou bastante disposta e eu acho que vai ser bem interessante poder disponibilizar esses dados, porque isso pode nos ajudar bastante a entender mais como funciona, ou o que que cada pesquisador tem feito, como que uma pesquisa se complementa com a outra, e assim, sucessivamente.

- Subcategoria “Financiamento público da pesquisa”

O argumento de que pesquisas que receberam financiamento público deveriam ser disponibilizadas abertamente, bem como os dados que as geraram, foi um dos fatores apontados por entrevistados (em especial, PS2, PS3 e PS8), no que se refere a estímulos à abertura de dados.

De acordo com PS8, a produção científica no Brasil é caracterizada, sobretudo, pela condução de pesquisas mediante financiamento público. Sendo assim, PS8 defendeu que dados produzidos sejam igualmente disponibilizados. Também sugeriu que isso ocorresse mediante o cumprimento de acordos nos quais o pesquisador compartilharia seus dados em alguma plataforma pública, após determinado período. Isso, para preservar o direito de o pesquisador explorá-los da maneira que almeja,

antes de sua total disponibilização. A esse respeito, PS8 explicou:

Por outro lado, eu acho que quando envolve dinheiro público, os financiadores deveriam pensar nisso, como que esses dados vão ser utilizados? Como aqui no Brasil você tem muito disso, do financiamento público da pesquisa, poderia ser uma contrapartida, você tem uma janela aí para manter o dado, depois, esse dado tem que ir para uma plataforma x aí qualquer, tem que ver onde toda e qualquer pessoa vai ter acesso a esses dados. Pelas características do Brasil, dessa relação da pesquisa com o dinheiro público e essa dependência que nós temos muito grande com o dinheiro público, esse seria, talvez, o caminho mais razoável de possibilitar esse acesso. Assim, teve financiamento, terminou a pesquisa, você tem dois anos para lidar com os dados, depois você tem que depositar eles em algum lugar, para que todos tenham acesso. Mas poderia funcionar tranquilamente.

Nesse mesmo contexto, PS2 comentou que seria muito natural o processo de abertura de dados, principalmente quando esses forem obtidos mediante financiamento público. Assim, PS2 explicou que gera informações úteis para o aperfeiçoamento de políticas públicas, sendo comum a publicação de conteúdos científicos resultantes de estudos financiados por recursos públicos:

Olha, isso meio que acontece ao natural, como eu te disse, pela obrigação, pela pressão que a gente tem por publicar, por tornar público os recursos que são feitos a destinação, fazer um bom artigo, um artigo que ajude. Como sempre eu digo, o nosso papel é gerar informação, o meu papel não é fazer política, o meu papel é gerar informação para que essas informações sirvam para aperfeiçoar políticas. Eu acho que minha obrigação é de fazer isso seguindo os canais naturais de quem é um cientista.

Ainda se referindo à abertura de dados de pesquisa, PS3 comentou que parte expressiva da área de Ciências humanas conduz pesquisas devido financiamento público recebido. Logo, pesquisadores teriam a obrigação de retornarem à sociedade aquilo que foi investido em pesquisas, mediante a publicação de artigos e dados.

- Subcategoria “Liderança institucional”

Com base nos dados obtidos no estudo, observou-se que a existência de uma liderança institucional condutora de diretrizes para a abertura de dados foi um dos fatores que estimulariam tal abertura dentre os entrevistados, principalmente PS2 e PS3.

Referindo-se a agências de fomento, PS3 comentou que essas não costumam requerer bancos de dados gerados por pesquisadores, após a conclusão de pesquisas financiadas. O que geralmente ocorre é a prestação de contas mediante a apresentação de relatórios. Entretanto, aspectos referentes a dados, tais como as estratégias de preservação de dados, não são pontos amplamente abordados. Por

essa razão, sugere-se a existência de uma liderança institucional que seja mais sensível à questão de dados de pesquisa.

Segundo PS2, faltam diretrizes específicas no que se refere à abertura de dados, principalmente, aqueles gerados por pesquisas financiadas por agências de fomento brasileiras. Por conseguinte, algumas indagações relacionadas a dados permanecem entre pesquisadores na área de Sociologia, tais como questionamentos relativos a direitos autorais e tratamento técnico de dados, que incluem a adoção de mecanismos de proteção aos mesmos. Sobre isso, PS2 comentou:

Eu não me sinto nem seguro para te dizer, porque eu escrevo um livro, como agora eu escrevi, e eu cedo os direitos autorais para a universidade. Uma cessão de direitos durante dez anos, eu não posso utilizar esse livro em outros recursos, é tudo da universidade. Agora, os meus dados, olha, que pergunta difícil, né? Muito difícil de responder e eu, sinceramente, não me sinto hoje. A maneira com que as coisas estão, tipo quem é que vai cuidar disso? As pessoas passam, as instituições ficam, mas enquanto isso, que pessoas são essas? Como a sociedade tem mecanismos para fazer com que esses dados sejam? É garantida a confidencialidade?

Uma liderança institucional no âmbito de dados de pesquisa também poderia contribuir para a conscientização de pesquisadores sobre a importância e benefícios trazidos pela abertura de dados. Considerando que muitos cientistas ainda trabalham de modo individual e, até mesmo competitivamente, PS4 valoriza a colaboração entre pesquisadores para o avanço da Ciência, conforme expôs em sua fala:

O que facilitaria era ter recursos para isso e ter uma campanha de conscientização dos pesquisadores da importância disso. E ainda existe uma parte grande dos pesquisadores que trabalham com a cabeça em código fechado e um jeito competitivo, sendo que ciência no mundo inteiro é colaboração. É com colaboração em pesquisa que a gente avança, mas aqui ainda tem muito isso, né? Do pequeno mundo e tal, na cabeça em código fechado.

Um outro aspecto abordado por PS8 foi a vaidade de pesquisadores e o apego desses em relação aos dados que geram, o que dificultaria a abertura de dados em um contexto mais amplo. Entretanto, PS8 destacou que tal situação deveria ser repensada quando se tratar de dados que foram obtidos mediante financiamento público, de modo que haja um equilíbrio entre o interesse particular do pesquisador e o interesse público:

Que você possa usufruir disso, mas que depois, toda população possa, ou todo pesquisador, ou quem quiser utilizar dos seus dados. Então, talvez a questão é até muito mais subjetiva desse apego aos dados nesse aspecto que eu acho que resvala na vaidade. Agora, por outro lado, eu acho que quando envolve dinheiro público, os financiadores deveriam pensar nisso, como que esses dados vão ser utilizados?

4.3.2.2 Categoria “Fatores inibidores à abertura de dados de pesquisa”

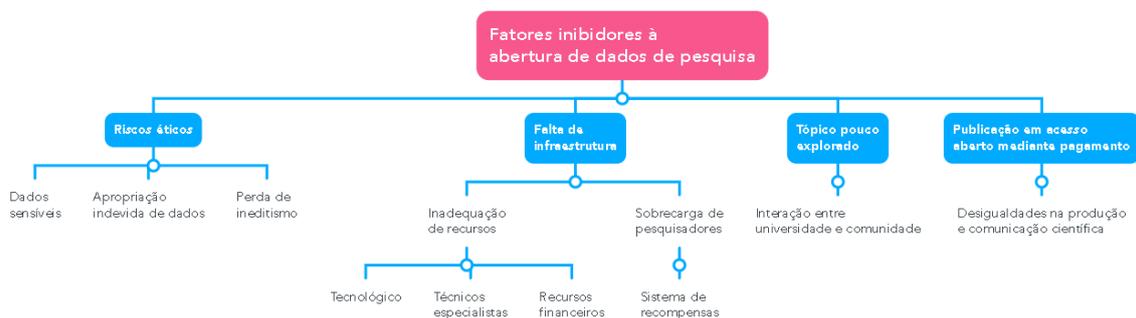
A categoria “Fatores inibidores à abertura de dados de pesquisa” representa os aspectos observados no estudo que contribuem para desestimular a abertura de dados de pesquisa. Nessa direção, trata-se do conjunto de forças que agem sobre pesquisadores, inibindo-os ou obstaculizando-os, no que se refere à abertura dos dados que produzem.

Como pode ser visto na Figura 42, esta categoria desenvolveu-se nos termos de quatro subcategorias, sendo elas: riscos éticos; falta de infraestrutura; tópico pouco explorado; publicação em acesso aberto mediante pagamento. A primeira subcategoria apontou que dados podem estar suscetíveis a violações éticas, o que inclui riscos relacionados à exposição de dados sensíveis, à apropriação indevida de dados e perda de ineditismo de publicações.

Por sua vez, a segunda subcategoria mencionada apontou a falta de infraestrutura necessária à abertura de dados como um fator inibidor para tal disponibilização. Nesse sentido, entrevistados relataram que os recursos (principalmente tecnológicos, de técnicos especializados e financeiros) destinados a essa finalidade são insuficientes. Além disso, muitos pesquisadores encontram-se sobrecarregados com tantas atividades administrativas as quais são responsáveis. Considerando que a disponibilização de dados abertos de pesquisa demanda tempo e trabalho de pesquisadores, foi sugerido dentre os entrevistados, que houvesse um sistema de recompensas que bonificasse o cientista que optar pela abertura dos dados que produz.

A terceira subcategoria caracteriza-se, sobretudo, por compreender a abertura de dados de pesquisa como um tópico ainda não explorado em sua totalidade, o que dificulta uma maior divulgação do tema e conseqüente interação entre comunidade científica e sociedade. Por fim, a subcategoria “Publicação em acesso aberto” representou os impasses, principalmente financeiros, para que pesquisadores publiquem artigos e dados em acesso aberto. Tais impasses refletem algumas desigualdades na produção e comunicação científica, no que se refere a condições financeiras para o pagamento de taxas de cobrança para a publicação.

Figura 42 – Categoria Fatores inibidores à abertura de dados de pesquisa - Sociologia



Fonte: elaborado pela autora

- Subcategoria “Riscos éticos”

Com base nas informações obtidas nos relatos dos entrevistados (principalmente PS5, PS6 e PS9) notou-se que tais pesquisadores lidam rotineiramente com dados sensíveis. Esses, em geral, se referem a situações, sentimentos e contextos bastante individuais vivenciados por participantes de pesquisas. Não são facilmente expostos, muitas vezes são sigilosos e, por conseguinte, necessitam de maior proteção ou não podem ser divulgados devido a acordos de confidencialidade estabelecidos entre pesquisador e participante da pesquisa.

Nessa direção, PS6 destacou que tem disposição para compartilhar dados de pesquisa, desde que não se trate de dados sensíveis. Isso, devido a exposição a riscos éticos pelos quais esses dados estariam suscetíveis após divulgação. A título de ilustração, PS6 disse que pode realizar estudos com indivíduos que estão na trajetória criminal e que, dependendo da situação, a exposição indevida de determinados fatos e também de outros indivíduos poderia até colocar vidas em risco. Sobre isso, PS6 explicou:

Eu só não compartilharia, por exemplo, entrevistas muito sensíveis, alguns atores que às vezes a gente, sei lá, que estão na trajetória criminal, qualquer coisa assim, que aí é mais delicado. Porque a gente pode colocar a vida dessa pessoa em risco, mas por exemplo, esse tipo de informação que eu estou falando contigo, né? Da produção de dados e mesmo entrevistas, resultados de grupos focais mais gerais, quanto material quali, quanto quanti, eu não vejo nenhum problema.

Semelhantemente, PS5 afirmou que nem todos os dados que produz poderiam ser disponibilizados, devido a alguns tratarem de temáticas sensíveis as quais, eticamente, não poderiam ser expostas. Portanto, PS5 costuma focar-se na

caracterização de processos e situações relacionados a pessoas, e não na descrição das mesmas, propriamente dizendo. Até porque, segundo PS5, participantes das pesquisas que realiza confiaram no acordo de que não teriam suas identidades reveladas.

Outro aspecto relacionado a riscos éticos no contexto de dados de pesquisa é a apropriação indevida desses, conforme apontou PS2. Expondo sua opinião, disse que determinados pesquisadores poderiam utilizar conjuntos de dados sem atribuir a verdadeira autoria do pesquisador que os produziu, visando a obtenção de vantagens no âmbito científico. Sendo assim, PS2 disse que construiu sua reputação científica a partir da sistematização rigorosa da produção de dados, não sendo um alpinista, o qual para obter *status* científico, retém para si o mérito de outra pessoa. No que se refere a tal aspecto, PS2 declarou:

Não tenho interesse nenhum, a não ser produzir conhecimento de qualidade. Sabe, construí minha reputação em cima de uma atitude sempre rigorosa, desconfiando até dos meus próprios dados e assim formar meus alunos com isso. Não tem lugar para alpinista, já tive brigas com colegas meus porque chamei de alpinista. Alpinista é aquele que para chegar no topo da montanha, se tiver que pisar em cima dos outros ele faz isso, e eu não construí a minha reputação em cima de alpinismo. Se tiver que chegar ao lugar onde eu cheguei, eu cheguei pelos meus méritos e não mentindo, e não [sendo] um impostor, um manipulador.

Nesse mesmo contexto, destaca-se a preocupação de pesquisadores como PS3, em relação à perda do ineditismo ocasionada pela publicação antecipada de dados gerados em pesquisas. Para tanto, PS8 sugere a determinação de um período para que responsáveis de uma pesquisa possam ter a exclusividade inicial para manusear os dados produzidos, de acordo com seus interesses, e só após o cumprimento do prazo, tais dados poderiam estar disponíveis abertamente. Sobre isso, PS8 comentou:

Eu não veria problema algum em compartilhar os dados da pesquisa que eu por ventura venha fazer. As que eu já fiz, eu nem sei se serviria, e essa, [pesquisa a qual está desenvolvendo] como está bem próxima, daria para recuperar, mas eu acho que tem essa questão também do uso, quem vai usar primeiro os dados e, às vezes, esse tempo para que os responsáveis pela pesquisa possam ter essa exclusividade inicial de trabalhar com os dados. É uma questão também importante para depois disponibilizar. Teria um prazo para isso e depois disponibilizaria essa pesquisa.

- Subcategoria “Falta de infraestrutura”

Os relatos dos entrevistados, principalmente os de PS4 e PS9, demonstraram que a falta de infraestrutura, muitas vezes ocasionada pela inadequação de recursos

destinados à área da pesquisa, constitui-se um fator inibidor à abertura de dados. De acordo com PS4, trata-se de um problema associado à produção científica no Brasil, onde não se observa uma gestão de projetos e financeira eficazes.

Ao relacionar o contexto universitário com a disponibilização de dados gerados por pesquisadores, PS4 relatou que seria favorável à instituição em que atua, caso a mesma se preparasse gerencialmente para viabilizar a abertura de dados. Nesse sentido, tal medida incluiria a arquitetura de banco de dados quali ou quantitativos, a disponibilização de especialistas na gestão de dados e o suporte de estrutura tecnológica para esse fim. Entretanto, essa ainda não é uma realidade vivenciada por PS4, que, por sua vez, tem notado uma demanda crescente a pesquisadores em universidades, inclusive, de cunho administrativo.

Ainda segundo PS4, no Brasil costuma-se adotar um modelo no qual primeiramente instauram-se demandas, para somente depois, desenvolver a infraestrutura necessária para isso. Por essa razão, muitos teriam desistido da carreira científica para exercerem outras carreiras. Os pontos até aqui mencionados foram explicados da seguinte maneira por PS4:

Agora, se falarem: 'olha, a partir de hoje, nós vamos disponibilizar projetos de montagem de bancos de dados qualitativos, quantitativos, etc. para serem disponibilizados em 2028' e durante esses cinco anos, nós vamos bancar profissionais e estrutura para que isso seja feito na universidade e aí em 2028, vai ser obrigatório e essa infraestrutura vai estar funcionando, maravilhoso! Entende? O problema é só esse, não dá para você demandar para o pesquisador uma coisa que você não oferece nenhuma condição dele fazer. Para você fazer alguma coisa em qualquer lugar do mundo, vai ser: 'oh, a gente quer que você sequencie o DNA', então, a gente vai ter que financiar uma máquina de sequenciamento de DNA, não dá para o cara fazer da cabeça dele. Então, o que a gente está acostumado a ver no Brasil é sempre esse modelo, primeiro vem a demanda e depois vem a infraestrutura, e quando vem, e na verdade não vem. Então, as pessoas se desenvolvem a um ponto em que isso aqui fica meio impossível, aí ou elas desistem de fazer carreira científica. Mesmo sendo professores, está cheio de colegas que desistem da carreira científica propriamente dentro da universidade, porque tem outras carreiras.

Considerando o contexto no qual pesquisadores se sentem sobrecarregados em relação a demandas acadêmicas e administrativas nas universidades as quais atuam, PS4 mencionou a criação de um sistema de recompensas como um aspecto que também estimularia a abertura de dados de pesquisa.

De acordo com PS4, seria interessante que o pesquisador que disponibilizasse abertamente os dados que produz obtivesse vantagens, tais como o aumento no valor concedido para o financiamento de projetos de pesquisa, a partir da

implementação de um sistema de pontuação que gratificasse cientistas que mais compartilhassem dados abertos.

Em um de seus comentários, PS4 expõe que pesquisadores geralmente encontram-se envolvidos com tantas atividades, que o compartilhamento de dados representaria mais uma atribuição a ser realizada, sem notáveis ganhos individuais. Nesse sentido, alguns pesquisadores não estariam dispostos a compartilharem dados abertos, conforme explicou:

A questão é que você não tem incentivo nenhum na carreira para que você possa fazer as coisas acontecerem direito. Aí, imagina se tem um pesquisador, que você começa a exigir dele mais um milhão de outras coisas: 'oh, agora você vai ter que passar pelos sete comitês de ética, agora você vai ter que passar por duzentos pareceristas, e depois, você tem que disponibilizar todos os seus dados'. O pesquisador fala: 'para que eu vou sofrer tanto, se eu vou ganhar a mesma coisa no final do mês? Melhor eu não fazer pesquisa', você entende?

Sob outro enfoque, PS2 comentou que, em razão da falta de infraestrutura, não há garantias de que a gestão de dados será adequada em universidades. Já PS9, destacou que tal gestão exige habilidades técnicas, as quais não dispõe neste momento, por não ter uma formação nessa área específica. Assim, exemplificou isso ao dizer que não tinha muita afinidade com determinados *softwares* que auxiliam a descrição e análise de dados quantitativos por considerá-los de sintaxe complexa. Logo, faz-se necessária formação técnica adequada para lidar com diferentes sintaxes relacionadas à produção, descrição e análise de dados de pesquisa.

- Subcategoria “Tópico pouco explorado”

A partir dos dados obtidos no estudo, percebeu-se principalmente nos relatos de PS2, PS9 e PS10, que o tópico da abertura de dados de pesquisa ainda é pouco explorado no contexto dos entrevistados. Relacionando o ambiente universitário e a sociedade como um todo, PS9 disse que a abertura de dados pode promover maior interação entre universidade e comunidade, uma vez que esse é um tema pouquíssimo discutido entre ambos.

Ainda segundo PS9, a universidade, muitas vezes, é reconhecida como um setor de filtro no qual a informação produzida permanece restrita aos integrantes desse setor. Logo, a divulgação científica poderia ser enfatizada e estendida ao

contexto da comunidade não acadêmica, de modo que o tópico da abertura de dados de pesquisa seja amplamente abordado.

Semelhantemente, PS10 disse que a abertura de dados de pesquisa é um tópico que dificilmente será implementado a curto prazo no contexto brasileiro, em razão de ter sido pouco discutido. Para PS10, esse não é um tema recorrente nas pautas do programa de pós-graduação o qual integra, como se pode observar em seu comentário:

Em um curto prazo, eu acho remoto. Como eu te falei, não é um tema debatido isso, é uma evidência de que não é um assunto premente, um assunto que está na ordem do dia. Então, eu não vejo assim que possa ter resistências, certamente vão ter discussões, tipo essa que eu coloquei aqui, comitê de ética, de sigilo em relação aos informantes, mas reafirmo que não é um tema que está colocado. Então, eu acho que em um curto prazo, há poucas chances que isso ocorra na universidade, pegando o meu microuniverso onde eu me encontro nesse momento.

Na fala de PS10, também se nota que a promoção de discussões sobre a abertura de dados é interessante para estimular tal abertura entre pesquisadores da área e sanar questionamentos a respeito de vantagens e implicações para a comunidade científica. Só então, seria possível a introdução de práticas de gestão de dados de pesquisa específicas.

Acho que uma questão que poderia ser estimulada é maior discussão dentro dos programas de pós-graduação, dentro dos institutos de pesquisa sobre isso, sobre essa ideia, essa noção de ciência aberta. [...] uma discussão, quais as vantagens? O porquê de se promover esse tipo de ideia, essa noção de pesquisa aberta, de ciência aberta. Me parece que passaria por aí, primeiro, discutir o assunto, depois, pensar no que você falou: planejamento, em técnicas de armazenamento de dados. Depois para uma parte instrumental, mas eu acho que falta ainda uma discussão mais difusa sobre o assunto.

- Subcategoria “Publicação de dados em acesso aberto mediante pagamento”

Os dados coletados no estudo demonstraram que a publicação de artigos e dados em acesso aberto mediante pagamento se constitui um aspecto inibidor para a abertura de dados. De acordo com PS8, o custo para a publicação de um artigo em acesso aberto cobrado por determinados periódicos gira em torno de dez mil dólares.

Nessa direção, PS8 continua explicando que no Brasil, muitos periódicos têm acesso gratuito e disponibilizam dados de maneira aberta. Entretanto, essa não é uma realidade comum no âmbito de revistas internacionais da área, nas quais observa-se a exigência de pagamento para a publicação. À luz do exposto, fica evidente o

alto custo de manutenção de tais revistas, relacionados principalmente com os processos de editoração envolvidos. Sobre os aspectos mencionados, PS8 relatou:

Aqui no Brasil, muitas revistas têm acesso gratuito, são abertas. Enfim, mas essa não é uma realidade no mundo. [...] então, muitas revistas brasileiras são abertas, você acessa ali gratuitamente, quantas vezes você quiser, baixa o artigo, mas revistas internacionais, isso não ocorre [...] mas é um ponto também que eu acho que tem que ser colocado, pagar para publicar, evidente que é um custo, a revista tem um custo de manutenção que deve ser caríssimo de editar, editoração, coisa assim, a manutenção do acesso, os funcionários, enfim, tem um custo, alguém tem que pagar por isso, é um ponto que eu acho que vale apenas discutir também.

Ainda segundo PS8, tal situação reflete-se em desigualdades no que se refere à comunicação científica, de modo a privilegiar aqueles que podem pagar pela publicação de artigos e dados. Por conseguinte, pesquisadores fora do eixo Estados Unidos - Europa ocidental estariam quase inviabilizados desse tipo de publicação, devido à insuficiência de recursos destinados a esse fim.

Nesse mesmo contexto, um outro aspecto foi mencionado: a seleção de conteúdos a serem publicados. Em nível nacional, PS8 sugeriu que autores do norte e nordeste tenham dificuldades para publicar tais conteúdos em revistas do sul e sudeste do país. Em nível internacional, tal situação se refletiria, igualmente, em relação a periódicos do sul e norte global, conforme explicou PS8:

Sem contar outros aspectos não materializados que envolvem a seleção ou não do que vai ser publicado, pensando em revistas internacionais. De repente, você pensando em revistas nacionais, você pode ter uma relação aí sul-sudeste com norte e nordeste, então, ao todo, essas pessoas do norte e nordeste tem dificuldade de publicar em revistas do sul e sudeste, aí você pode passar isso para a questão do sul global, do norte global, essa dificuldade, enfim.

De outro vértice, o estudo observou que a proposição de obrigatoriedade de depósito de dados abertos de pesquisa estabelecida por instituições como universidades, agências de fomento e por editores científicos a pesquisadores ora foi entendida como um fator que poderia estimular tal abertura, ora como um fator que poderia inibi-la. Em razão desse fato, no presente trabalho se optou pela criação de duas subcategorias que compilam cada uma das opiniões e argumentos favoráveis ou desfavoráveis à referida obrigatoriedade.

Nessa direção, a opinião favorável à obrigatoriedade de depósito de dados como instrumento de estímulo para a abertura de dados de pesquisa está representada na categoria: “Favorável à obrigatoriedade”. Por outro lado, a opinião

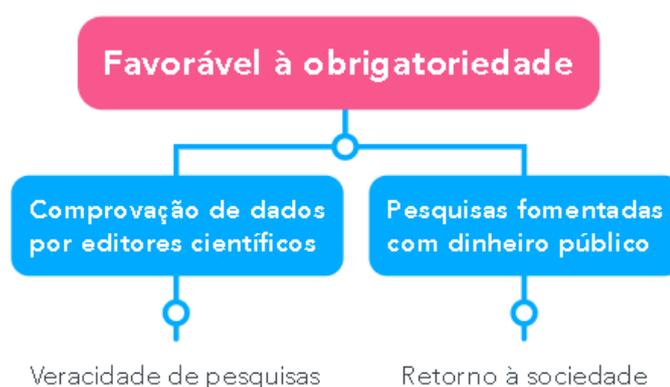
desfavorável a essa obrigatoriedade, de modo a torná-la um aspecto inibidor à abertura de dados de pesquisa, está representada na categoria: “Desfavorável à obrigatoriedade”.

4.3.2.3 Categoria “Favorável à obrigatoriedade”

Como observado na Figura 43, a categoria “Favorável à obrigatoriedade” pode ser entendida nos termos da subcategoria “Comprovação de dados por editores científicos” e da subcategoria “Pesquisas fomentadas com dinheiro público”. A primeira subcategoria refere-se ao argumento de que o acesso a dados que embasaram determinada pesquisa é fundamental para atestá-la quanto a sua veracidade.

Por seu turno, a segunda subcategoria ampara-se no argumento de que pesquisas realizadas mediante o recebimento de recursos públicos devem ter os dados publicamente disseminados. Nesse sentido, a sociedade receberia um retorno sobre o financiamento que concedeu a pesquisas, por meio de suas diversas instituições.

Figura 43 – Categoria Favorável à obrigatoriedade - Sociologia



Fonte: elaborado pela autora

- Subcategoria “Comprovação de dados por editores científicos”

Dentre as opiniões de entrevistados favoráveis à obrigatoriedade da disponibilização de dados abertos de pesquisa, destacam-se a de PS2, PS3 e PS8.

Isso, principalmente, pela possibilidade de comprovação de dados, ou seja, trata-se de mais um mecanismo útil para o ateste da veracidade de pesquisas realizadas.

Nessa direção, PS2 comentou que se o editor contesta algum aspecto a respeito de dados, o pesquisador tem o dever de demonstrar como esses foram coletados, e, por conseguinte, que não é um impostor em relação ao conteúdo a ser publicado. Sobre a relevância da verificação da fidedignidade de informações relativas aos dados que produz, PS2 comentou:

Eu acho que tu tens a obrigação de mostrar o que os teus dados têm, e que aquilo não é impostura. Eu acho que é uma atitude lógica de quem vive em um ambiente em que todo momento tu és questionado, criticado, desafiado. Muitas vezes, as pessoas não concordam com os teus dados, e podem dizer 'olha, eu acho que esses dados aqui, não!' Eu digo: 'poxa, vida! Tá, se eles querem que eu provo, eu provo! [...] eu sou do tempo em que você ia para uma banca e ia assim, com as tuas entrevistas transcritas, apresentava os dados e aí o cara dizia: 'só um pouquinho, isso é verdade'? Não, tem aqui, oh! Só me dá um tempo que eu vou localizar a entrevista que está aqui que a informação está colocada e tal'.

De acordo com PS3, cada vez mais, observa-se que editores científicos têm requisitado conjuntos de dados para replicar resultados, bem como para compreender melhor a dinâmica de técnicas adotadas em determinado estudo. Sendo assim, PS3 recordou que em dada oportunidade, um editor de um periódico lhe solicitou o envio dos dados da codificação de um estudo que havia realizado, e que tal procedimento foi fundamental para a replicabilidade de processos analíticos envolvidos na pesquisa.

A esse respeito, PS3 explicou:

Os editores científicos, eles têm solicitado, cada vez, mais os bancos de dados. Inclusive, recentemente, eu passei por um processo desse e era uma análise qualitativa e a gente tinha feito tudo em português, mas a publicação era em inglês. E aí, o editor pediu que a gente mandasse a codificação em português, não interessava o problema da língua porque ele queria ver como tinha sido feita essa codificação. Então, eu acho que, especialmente do ponto de vista, não sei dos dados, mas do ponto de vista da replicabilidade, porque ciência é exatamente isso, a possibilidade de você replicar aquela análise e ver se você chegou na mesma conclusão, eu acho que isso tem crescido, inclusive, do ponto de vista de importância. Eu acho que os editores têm um papel bem importante nesse sentido.

Nesse contexto, destaca-se que a obrigatoriedade da disponibilização de dados de pesquisa exigida por editores científicos é recorrente no âmbito internacional. No Brasil, alguns periódicos já têm adotado essa medida. Segundo PS8, esse é um procedimento fundamental em casos nos quais suspeita-se da veracidade de determinados dados e das conclusões obtidas com base em resultados de pesquisa. Sendo assim, PS8 comentou:

Então, isso é necessário caso alguém levante uma suspeita: ‘não, eu acho esse resultado aqui não esperado, muito estranho e, se isso for verdade, ele vai romper com os paradigmas, aí, então, vamos ver se foi isso mesmo que foi encontrado, enfim’. É importante, eu acho que as revistas brasileiras deveriam fazer isso também, é um padrão que você começa a observar nas revistas internacionais. Eu fico pensando na minha área, as revistas que eu tive contato, elas estão pedindo isso, até me perguntam: ‘esse dado, ele é público ou não? É público, por exemplo, onde que as pessoas podem acessar esses dados se ela quiser replicar seus estudos?’ [...] isso eu acho que é importante no Brasil também. Isso eu acho que na área das ciências sociais está começando a ser exigido e eu acho uma boa sim, eu não vejo negativamente, não.

- Subcategoria “Pesquisas fomentadas por dinheiro público”

Alguns entrevistados favoráveis à obrigatoriedade da disponibilização de dados abertos exigida, principalmente, por universidades e agências de fomento pautaram-se no argumento de que pesquisas fomentadas por dinheiro público deveriam estar disponíveis à ampla audiência. Conforme disse PS3, estudos produzidos a partir de financiamento público deveriam ser retornados de alguma forma à sociedade:

Eu acho que me parece muito importante, mesmo porque na área das ciências humanas, boa parte desses estudos é produzido com financiamento público, então, a gente deveria ter a obrigação de retornar para a sociedade aquilo que a própria sociedade pagou a partir dos seus impostos para que a gente produzisse.

Referindo-se ao contexto da obrigatoriedade do depósito de dados abertos que poderia ser instituída por agências de fomento, PS8 demonstrou ser favorável a tal medida, desde que esse ato seja integralmente legal. Nessa perspectiva, a agência que financiou determinada pesquisa poderia requerer a disponibilização dos dados de pesquisadores, conforme o cumprimento de termos de liberação de dados firmados entre agência e pesquisadores. Sobre isso, PS8 explicou:

Normalmente, o dinheiro vem dessas agências de fomento. De repente, elas poderiam (evidentemente consultando a legalidade disso, a constitucionalidade disso) impor: ‘não, tudo bem, você aceitou!’ Você tem um financiamento liberado, aceita o termo ali e nesse termo tem: ‘olha, depois de tanto tempo, os dados da pesquisa têm que ser liberados’. Como eu falei, é evidente que você vai ter que pensar nisso. Dados genéticos vão ser liberados? Como que vai ser? Que tipo de dados? Qual o nível de segurança?

Como se pode observar na fala de PS8, os termos estabelecidos entre agência de fomento e pesquisadores devem estar objetivamente declarados. Isso inclui a descrição dos dados que podem ser disponibilizados, por quanto tempo podem ficar tutelados por quem os gerou, quais os níveis de proteção para acessá-los, entre outros aspectos.

O argumento de que dados cuja produção foi financiada por instituições públicas deveriam ser disponíveis livremente, ainda que mediante determinação obrigatória, dialoga em nível direto com o aspecto no qual a sociedade se beneficia com o conhecimento científico. Trata-se de uma devolutiva à mesma, na qual presta-se contas de investimentos destinados à produção da ciência. Nessa direção, PS9 relatou que a disponibilização de dados abertos proporcionaria maior interação entre universidade e sociedade.

4.3.2.4 Categoria “Desfavorável à obrigatoriedade”

Conforme mostrado na Figura 44, a categoria “Desfavorável à obrigatoriedade” pode ser compreendida por intermédio da subcategoria “Livre-arbítrio do pesquisador”. Essa subcategoria apontou para o aspecto de que pesquisadores teriam a escolha livre e esclarecida sobre disponibilizar, ou não, dados que produziram. Assim, muitos pesquisadores teriam o interesse de explorar os dados que geraram em sua totalidade, não sendo interessante que fossem obrigados a disponibilizá-los a qualquer tempo.

Nesse contexto, destaca-se uma possível resistência de pesquisadores a decisões que não foram democraticamente construídas, tal como a obrigatoriedade de disponibilização de dados de pesquisa. Assim sendo, entrevistados sugeriram o estabelecimento de diálogos, principalmente entre pesquisadores, instituições de ensino e pesquisa, e agências de fomento.

Figura 44 – Categoria Desfavorável à obrigatoriedade - Sociologia



Fonte: elaborado pela autora

- Subcategoria “Livre-arbítrio do pesquisador”

Com base nas informações obtidas no estudo, notou-se que alguns entrevistados, principalmente PS5, PS6 e PS7, declaram-se contra uma possível obrigatoriedade em relação à disponibilização de dados abertos gerados por pesquisadores. Nessa direção, o principal argumento exposto foi o de que o pesquisador deve ter o direito de escolher se vai disponibilizar seus dados, ou não.

Segundo PS5, pesquisadores têm o livre arbítrio para definir o destino dos dados que produzem, não cabendo a ideia de obrigatoriedade instituída por universidades, agências de fomento ou editores científicos. Além disso, comentou que um pesquisador poderia ter planos de usar dados produzidos por um determinado período, não sendo conveniente disponibilizá-los antes de explorá-los em sua totalidade.

A título de ilustração, PS5 relatou que, uma vez, disponibilizou dados de pesquisa em um periódico, por se tratar de um estudo específico o qual não incorreria em riscos éticos, além de poder contribuir para o avanço científico na área. Destacou-se, nesse contexto, que a disponibilização de dados ocorreu mediante sugestão, e não condição, para a publicação do artigo vinculado aos dados mencionados. Sobre isso, PS8 falou:

A revista que eu publiquei [dados de pesquisa] foi essa experiência daí que foi muito interessante. Ela não colocou como condição, ela colocou como um avanço na ciência, tudo isso. Aí, a gente meditou bem, aí a gente falou: 'oh, não vai ter problema, vamos contribuir com esse avanço da ciência, mas, outra pesquisa, eu não faria isso.'

Contra-pondo-se a uma possível obrigatoriedade de disponibilização de dados abertos de pesquisa instituída pela universidade em que atua e por agências de fomento, PS6 destacou uma provável resistência por parte de pesquisadores da área. Isso devido a uma tradição brasileira na qual, primeiramente, um projeto se torna obrigatório, para só então ser amplamente discutido.

Com relação à abertura de dados no contexto brasileiro, PS6 defendeu um caminho de construção caracterizado, sobretudo, pelo diálogo entre pesquisadores, instituições de ensino, pesquisa e agências de fomento. Ademais, é de extrema importância que pesquisadores passem a conhecer melhor as bases epistemológicas

da Ciência aberta por meio de linhas de convencimento, e não devido o cumprimento de determinada obrigação.

Semelhantemente, PS7 alegou que qualquer obrigatoriedade em relação a dados de pesquisa é uma atitude arbitrária, uma vez que se trata de um tema complexo que deveria ser democraticamente construído. Nesse sentido, defendeu que o produtor de dados deveria escolher disponibilizá-los abertamente, ou não, até porque, muitos desses encontram-se no estágio parcial da pesquisa, não podendo ser expostos até para não incorrer em futuros erros de interpretação e análise.

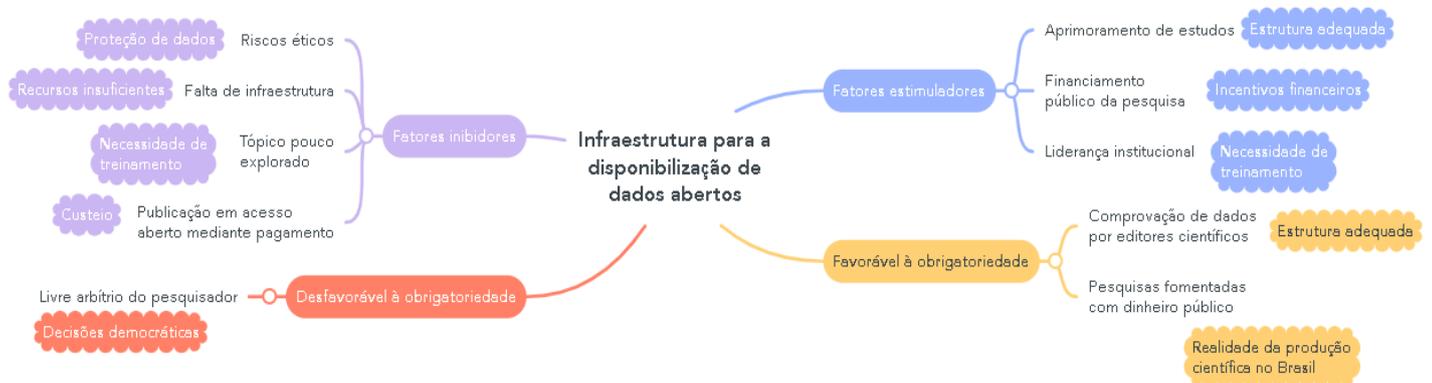
Ainda segundo PS7, agências de fomento até poderiam instituir tal obrigatoriedade como condição para a distribuição de recursos em projetos de pesquisa aprovados. Entretanto, comentou que haveria uma possibilidade muito grande de rejeição da comunidade acadêmica brasileira, que estaria avessa a essa arbitrariedade. Isso porque, para PS7, a disponibilização de dados abertos deveria ser uma adesão voluntária de pesquisadores, e não uma obrigação.

Eu acho uma coisa muito ruim, mas eu acho que as agências de fomento, talvez, infelizmente, teriam mais forças para fazer isso do que as próprias universidades, porque elas poderiam estabelecer como regra para a aprovação dos projetos, e isso seria muito ruim. Isso daria um quiproquó [problema, resistência] gigantesco. Porque a comunidade não só da [nome da universidade em que atua], mas assim, a comunidade acadêmica do Brasil cairia matando, tipo, seria muito avesso a esse tipo de posicionamento. Se eu entendi, foi um posicionamento arbitrário, né? [...] eu acho importante a gente pensar em dados abertos, mas como eu disse, tem que ser uma coisa democraticamente construída, não imposta e que tenha e que parta também de uma adesão voluntária do pesquisador.

4.3.3 Codificação seletiva e teoria substantiva referente aos fatores que influenciam a abertura de dados de pesquisa em Sociologia

No que diz respeito à área da Sociologia, dados obtidos no estudo constataram a centralidade da categoria “Infraestrutura para a disponibilização de dados abertos” nas relações estabelecidas com as demais categorias observadas. Assim como é representado na Figura 45, pode-se destacar tais relações da seguinte forma:

Figura 45 – Infraestrutura para a disponibilização de dados abertos e sua relação com fatores que influenciam a abertura de dados - Sociologia



Fonte – elaborada pela autora

- para que dados abertos de pesquisa se tornem insumos para o aprimoramento de estudos na área de Sociologia e promovam a comprovação de resultados por meio da veracidade de informações que veiculam, é necessário que cientistas tenham a seu dispor infraestrutura adequada que viabilize tal abertura;
- a infraestrutura para a abertura de dados de pesquisa é indispensável para garantir a proteção de dados quanto a aspectos éticos (relacionados à exposição indevida de dados) e para viabilizar a publicação de dados em acesso aberto mediante pagamento, caso seja necessário;
- em geral, pesquisadores da área de Sociologia não dispõem de recursos suficientes para o desenvolvimento de seus estudos, refletindo na falta de infraestrutura adequada e disponível para a abertura de dados. Incentivos financeiros nessa direção estimulariam tal abertura;

- considerando que a produção científica brasileira caracteriza-se, sobretudo, pelo financiamento público para a sua execução, atualmente, verifica-se uma redução de recursos previstos em orçamentos públicos. Tal redução contribui para a falta de estrutura na produção e disponibilização de dados de pesquisa, que, por sua vez, requerem investimento tecnológico, financeiro e de pessoal;
- a falta de liderança institucional na área de Sociologia é notória para a gestão de dados de pesquisa e consequente abertura dos mesmos, evidenciando a necessidade de realização de treinamentos sobre o tema (em razão de se tratar de um assunto ainda pouco explorado);
- apesar da possibilidade de haver infraestrutura adequada à abertura de dados, a eventual obrigatoriedade de disponibilização dos mesmos estabelecida por instituições acadêmicas e científicas a pesquisadores ainda encontra resistência na área de Sociologia. Nessa, preza-se por decisões democraticamente construídas que não infrinjam o direito de livre-arbítrio dos pesquisadores.

Com base em estudos da literatura científica, a seção seguinte abordará a categoria central – Infraestrutura para a disponibilização de dados abertos – conforme as relações estabelecidas com as demais categorias identificadas no estudo, que, por sua vez, se referem a fatores que influenciam a abertura de dados de pesquisa.

4.3.3.1 Infraestrutura para a disponibilização de dados abertos e sua relação com a fatores que influenciam a abertura de dados

O intercâmbio de dados contribui para o aprimoramento de estudos na área de Sociologia, tornando-a mais robusta em nível teórico. Esse benefício trazido essencialmente pelos dados, também ocorre em outras áreas, tal como pode ser ilustrado com a criação do repositório *COVID-19 Data Sharing/BR*⁷⁹, fruto de parcerias estabelecidas entre Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), Universidade de São Paulo (USP), Instituto Fleury, Hospital Sírio-Libanês e Hospital Israelita Albert Einstein.

79 Disponível em: <https://repositoriodatasharingfapesp.uspdigital.usp.br/>

Considerando o contexto pandêmico ocasionado pela doença Covid-19, destacam-se algumas iniciativas relacionadas a dados abertos que contribuíram para o aprimoramento de estudos nesse tópico de pesquisa e no desenvolvimento de novas investigações científicas. É o caso do repositório *COVID-19 Data Sharing/BR*⁸⁰, que disponibiliza em sua plataforma dados demográficos, clínicos e laboratoriais referentes ao estado de saúde de pacientes, além do número de internações e óbitos contabilizados.

No referido contexto, nota-se o quanto dados de pesquisa podem ser úteis à sociedade, principalmente, quando apresentam qualidade e inspiram confiança pela veracidade das informações que veiculam. Isso facilita a decisão de um pesquisador que pretende reutilizar determinado conjuntos de dados.

Por meio do processo de revisão por pares realizado em trabalhos submetidos a editores científicos, a produção científica a ser publicada é validada. Nesse sentido, a exigência da disponibilização de dados abertos a pesquisadores estabelecida por editores pode ampliar os meios de comprovação de dados, e, conseqüentemente, facilitar o ateste da veracidade de pesquisas realizadas.

Entretanto, conforme explicou Costa (2006), permanece a influência de um modelo de publicação científica com viés comercial. Autores recebem financiamento público para iniciar suas pesquisas e, após concluí-las, submetem manuscritos a editores científicos com os resultados obtidos. Quando o manuscrito é aceito, posteriormente, o mesmo é publicado pelos editores. Todavia, o que se verifica é que, em geral, o acesso ao trabalho do autor é mediado pela cobrança do custeio da assinatura do periódico no qual o estudo foi publicado.

Costa (2006) chama esse modelo de perverso e destaca que universidades e agências de fomento financiam pesquisas científicas, porém, resultados publicados, em geral, não se encontram publicamente acessíveis, nem mesmo a autores. Ao contrário, editores comerciais cobram taxas exorbitantes e impõem limitações de acesso a publicações. Sendo assim, Costa (2006) destaca o princípio no qual pesquisas financiadas por recursos públicos deveriam ser acessíveis à ampla audiência.

80 Disponível em: <https://repositoriodatasharingfapesp.uspdigital.usp.br/>

Dentre as escolas de pensamento sobre a Ciência Aberta demonstrada na literatura científica, Fecher e Friesike (2014) destacam Escola Pública, caracterizada, sobretudo, pela ênfase na necessidade de a ciência tornar-se mais acessível. A Escola Pública considera que tecnologias decorrentes da *Web 2.0* oportunizam a abertura do processo de pesquisa e o compartilhamento de informações a públicos não-especialistas.

Para que seja possível ampliar o acesso a dados abertos, e, por conseguinte, proporcionar o aprimoramento da área de Sociologia e facilitar a comprovação de estudos por meio da análise da veracidade de dados, é necessário que o pesquisador tenha a seu dispor infraestrutura suficiente que viabilize a abertura de dados. Neste estudo, tal infraestrutura foi mencionada por entrevistados como sendo o principal fator que influencia a abertura de dados, principalmente, no que se refere aos seguintes aspectos:

- **Proteção de dados em relação à ética** – para mitigar riscos éticos relacionados à exposição indevida de dados (principalmente aqueles de temáticas sensíveis), esses necessitam passar por tratamento técnico. Conforme descrito no modelo do ciclo de vida de dados propostos por *UK Data Archive*⁸¹, o processamento e análise de dados abrangem procedimentos de tratamento técnico nos quais dados de pesquisa são digitalizados, transcritos, traduzidos e anonimizados;
- **publicação de dados em acesso aberto mediante pagamento** – considerando esse tipo de publicação como um aspecto inibidor à abertura de dados, tal como verificou-se no presente estudo, nota-se que o custo para a publicação de artigos sob esses parâmetros pode chegar a milhares de dólares. Conseqüentemente, muitos pesquisadores entrevistados afirmaram que não possuíam recursos financeiros suficientes para arcar com essa cobrança.

Mais uma vez, toma-se nota da importância de o pesquisador ter disponível infraestrutura adequada para o desenvolvimento de pesquisas e gestão de dados, principalmente, no que se refere à abertura dos mesmos. Contudo, com base nos

81 Disponível em: <https://www.youtube.com/watch?v=-wjFMMQD3UA>

relatos dos entrevistados neste estudo, o que salta aos olhos é a falta de infraestrutura ocasionada, frequentemente, pela quantidade insuficiente de recursos destinados à atividade de pesquisa científica, e que se apresenta como o principal fator inibidor à abertura de dados.

A produção científica no Brasil caracteriza-se, sobretudo, pelo desenvolvimento de pesquisas subsidiadas pelo financiamento público. Isso é tão marcante no país, que dá indícios de que há uma dependência da produção científica em relação a recursos originários de instituições públicas. Por conseguinte, na medida em que há redução desses recursos previstos em orçamentos públicos, há prejuízos no contexto da ciência e da educação.

Com base no exposto, é interessante destacar algumas observações dispostas em relatório do *Observatório do Conhecimento sobre o balanço orçamentário brasileiro do ano de 2021*, elaborado pelos economistas Silva e Zelesco (2022). O Orçamento do conhecimento teve origem na necessidade de descrever como tem sido realizada a destinação de recursos públicos no Brasil, em relação à produção de conhecimento. Trata-se de um assunto muito importante, uma vez que análises a informações orçamentárias são úteis para complementar a avaliação de políticas públicas no país. Ademais, os dados do Orçamento do conhecimento apresentam valores deflacionados pelo Índice Nacional de Preços ao Consumidor Amplo (IPCA), tendo como referência o mês de março de 2022.

Dentre os resultados obtidos no Orçamento do conhecimento (2022) destaca-se que:

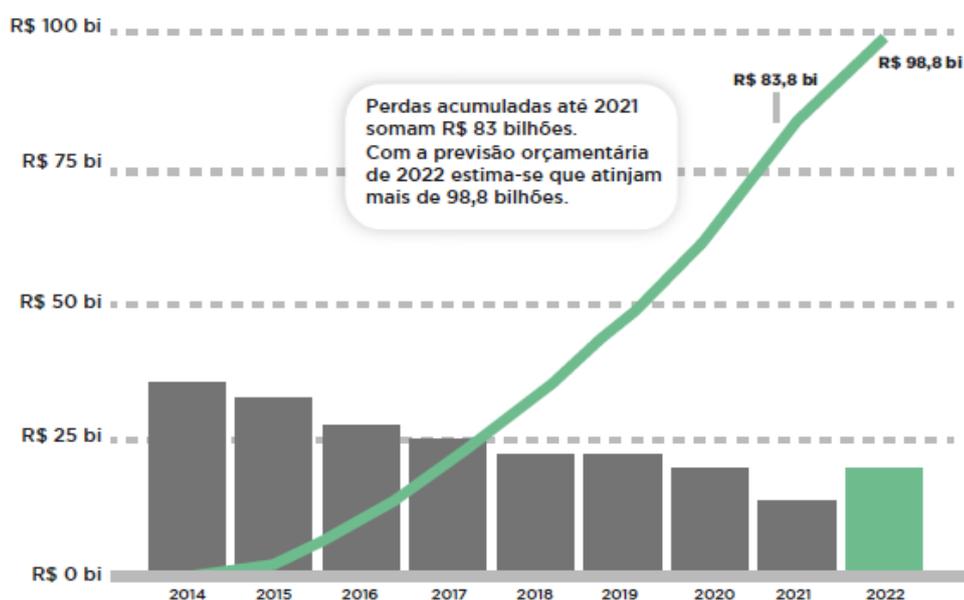
- em 2014; o Orçamento do Conhecimento representava R\$ 27,81 bilhões, passando para R\$ 10,57 bilhões em 2021;
- o Orçamento do Conhecimento acumula perdas estimadas de R\$ 83,8 bilhões entre os anos 2014 a 2021;
- A partir da dotação inicial, há previsões de que em 2022, o valor de R\$ 83,8 bilhões ultrapasse R\$ 99 bilhões, isso quer dizer que se o valor proposto em 2014 no Orçamento do Conhecimento tivesse permanecido o mesmo em termos reais, deveriam ter sido aplicados quase R\$ 84 bilhões a mais nas áreas relacionadas ao Ensino Superior, Pesquisa e Inovação. Para ilustrar tal situação,

considere-se que todos os gastos da Universidade Federal do Rio de Janeiro, em 2021, somaram aproximadamente R\$ 3,3 bilhões. Esse valor poderia manter 665.230 bolsistas de mestrado durante todos os 12 meses de 7 anos consecutivos;

- Considerando o percentual da dotação inicial que consta na Lei Orçamentária aprovada em relação aos valores liquidados, houve leve redução, dando indícios de que o Executivo federal vem executando, cada vez menos, valores aprovados no Legislativo;

O Gráfico 4 apresenta, em unidade de bilhões, as perdas acumuladas no Orçamento do Conhecimento, entre 2014 e 2021, além da estimativa desses valores para 2022.

Gráfico 4 - Perdas acumuladas no Orçamento do Conhecimento entre 2014 e 2022



Fonte: SIOP *apud* SILVA; ZELESCO (2022)

Como se pode observar no Orçamento do conhecimento (2022), as perdas orçamentárias impactam diretamente a produção científica no país, em especial, no ambiente acadêmico. Isso porque as universidades brasileiras exercem um papel de centralidade no ciclo produtivo da pesquisa científica, sendo diretamente afetadas pela diminuição de investimentos nas áreas de ciência e educação.

O relatório *Research in Brazil: Funding excellence Analysis prepared on behalf of CAPES by the Web of Science Group* (2019) descreve o panorama da pesquisa brasileira no contexto global, mediante a análise bibliométrica de dados sobre artigos brasileiros publicados no período de 2013 a 2018. No relatório, é possível perceber que quinze universidades representam mais de 50% de toda a produção brasileira nas áreas de Ciências da Saúde, Ciências Biológicas, Ciências Exatas e da Natureza, Ciências Agrícolas e Engenharia indexada na base de dados *Web of Science*, entre 2013 e 2018.

Ademais, apenas a Universidade de São Paulo (USP) produz mais que o dobro do número de publicações produzidas pela Universidade Estadual Paulista (UNESP), que é a segunda colocada neste *ranking* da produção científica. Destaca-se, ainda, que as quinze universidades mencionadas são públicas (WEB OF SCIENCE GROUP, 2019).

Logo, infere-se que a redução de recursos investidos em pesquisas brasileiras abre margem para a falta de estrutura na produção científica, bem como para a disponibilização de resultados de estudos e dados de pesquisa, uma vez que tal disponibilização é viabilizada por infraestrutura tecnológica, financeira e de pessoal adequada e suficiente.

Associada a esse fator que, predominantemente, influencia a abertura de dados, está a aparente falta de uma liderança institucional no que se refere à gestão de dados de pesquisa e à realização de treinamentos sobre o tema (por se tratar de um assunto ainda pouco explorado por pesquisadores e instituições de ensino e pesquisa).

Quanto a iniciativas voltadas a treinamentos acerca da Ciência aberta, pode-se destacar o projeto *Foster Plus*⁸² (acrônimo para *Facilitate Open Science Training for European Research*), que visa contribuir para a mudança do comportamento de pesquisadores por meio de três objetivos: 1) incentivar uma transformação cultural em favor da Ciência aberta no âmbito da produção científica; 2) consolidar uma rede

82 Disponível em: <https://www.fosteropenscience.eu/about#theprevproject>

de treinamentos a respeito da Ciência aberta; 3) fortalecer o ciclo de capacitações sobre Ciência aberta.

Na área de Ciências Sociais, pode-se destacar a condução de treinamentos sobre acesso, gestão e análise de dados os quais têm sido realizados pelo *Inter-university Consortium for Political and Social Research* (ICPSR). Além da capacitação, o consórcio promove o arquivamento de dados em diversas disciplinas, tais como Ciência Política, Economia, Sociologia e História.

No mesmo sentido, Abduldayan *et al.* (2021) recomendam que bibliotecas universitárias estimulem discussões a respeito da gestão de dados de pesquisa. Para isso, ressaltam a importância da realização de eventos como *workshops* em universidades, além da elaboração de políticas que norteiem o processo de submissão de dados em repositórios institucionais. Complementarmente, Aydinoglu, Dogan e Taskin (2017) sugerem a integração de cursos de ciência de dados ao currículo de departamentos acadêmicos e a realização de seminários sobre o tema.

Van Tuyl e Michalek (2015) ilustraram a capacitação de sobre gestão de dados na prática, quando foram realizados na *Carnegie Mellon University* (CMU), em 2012, *workshops* e discussões com pesquisadores a respeito do compartilhamento de dados em repositórios. Essa iniciativa resultou em um diagnóstico sobre infraestruturas necessárias para a criação de serviços de dados produzidos na universidade e indicou que a mesma teria que aprimorar muito de sua estrutura para que a gestão de dados ocorra efetivamente.

Com base no exposto, nota-se que ter à disposição infraestrutura de capacitação adequada para que pesquisadores conduzam a abertura de dados é fundamental no processo de decisão de cientistas por disponibilizarem abertamente os dados que produzem, ou não. Isso porque, apesar de uma possível obrigatoriedade de disponibilização de dados abertos estabelecida por instituições acadêmicas e científicas a pesquisadores, o presente estudo observou a tendência de uma resistência na área de Sociologia a decisões que não foram democraticamente construídas e que adentram o direito de livre-arbítrio dos pesquisadores. Portanto, não basta apenas argumentar que dados produzidos mediante financiamento público estejam, obrigatoriamente, disponíveis de maneira aberta.

4.3.4 Codificação aberta – Química

Esta seção apresentará a síntese dos principais fatores que influenciam a abertura de dados apontados pelos pesquisadores de Química entrevistados no estudo. Abordará as dimensões, categorias e propriedades identificadas por meio do processo de codificação aberta.

Convém relembrar que para garantir a anonimização de pesquisadores de Química entrevistados, quando os mesmos forem mencionados no texto, serão citados com o código “PQun^o”, em que “P” refere-se ao termo “pesquisador”, “Qu” ao termo “Química” e “n^o” ao número sequencial atribuído a cada pesquisador, de acordo com a ordem de realização das entrevistas.

4.3.4.1 Codificação aberta da entrevista com PQu1

PQu1 afirmou que tem disposição para compartilhar dados de pesquisas que produz, principalmente por se tratar de dados obtidos mediante financiamento público. Por essa razão, esses deveriam retornar à sociedade e proporcionar a geração de novos conhecimentos científicos.

Ao se referir a fatores que estimulariam a abertura de dados de pesquisa, PQu1 mencionou que exigências editoriais seriam fundamentais para esse fim. Para PQu1, muitos pesquisadores disponibilizam arquivos com conteúdos textuais, porém, dados de pesquisa não costumam ser publicados. Consequentemente, inúmeras coleções de dados encontram-se inacessíveis, ainda que existam mecanismos de inteligência artificial que viabilizem a recuperação da informação, como se pode observar em seu relato:

Eu vejo que a maioria das pessoas simplesmente prepara os PDFs a partir dos dados originais, mas não há o acesso a esses dados, porque... é uma coisa que eu acho que deveria ser muito mais comentada e talvez no final, quase exigida pelas editoras, porque é muitíssima informação que você está perdendo [...] hoje em dia, essa informação tem muito mais valor porque você tem tudo, toda a área de inteligência artificial, aprendizagem de máquina e você pode fazer muitas coisas com grandes coleções de dados.

Quanto a fatores que inibiriam a abertura de dados de pesquisa, PQu1 citou o aspecto do extremo zelo de geradores de dados, que dificulta a disponibilidade pública desses. Sobre isso, PQu1 explicou:

O que inibe é, às vezes, a falta de praticabilidade e, às vezes, tem grupos que pela sua natureza são bastante zelosos com seus dados. Então, simplesmente, não divulgam, ou se divulgam, é da mínima maneira possível.

Ao mencionar seu posicionamento em relação à possibilidade da universidade em que atua tornar obrigatório o depósito de dados que produz, PQu1 destacou a necessidade de infraestrutura tecnológica, física e técnica para lidar com o tratamento de dados, o que se constitui um enorme desafio na atual conjuntura acadêmica. A respeito disso, PQu1 propôs o depósito de dados atrelado à publicação de trabalhos como dissertações e teses. Ademais, pontuou alguns desafios de infraestrutura, como: disponibilidade de sistemas ou repositório de dados; proteção de dados; recursos financeiros; capacidade de preservação de dados.

Ao dar a sua opinião em relação à possibilidade de agências de fomento tornar obrigatório o depósito de dados, PQu1 explicou que diferenças disciplinares influenciam tal obrigatoriedade, uma vez que generalizações são descabidas em contextos disciplinares diferenciados. A título de exemplificação, PQu1 disse:

Então, o que é fácil, por exemplo, na química que no meu caso, é a área de ressonância magnética, pode ser complicado para químicos medicinais. Então, digamos que uma mesma regra para todos, às vezes é complicado.

Quanto à possibilidade de editores científicos tornarem obrigatório o depósito de dados, PQu1 concordou com tal obrigatoriedade, de modo que houvesse uma política mais incisiva sobre o compartilhamento e disponibilização de dados, bem como maior conscientização sobre o assunto por parte de pesquisadores.

O Quadro 58 apresenta as dimensões, categorias e propriedades identificadas na codificação aberta, a partir da entrevista de PQu1.

Quadro 58 – Codificação aberta da entrevista com PQu1: fatores que influenciam a abertura de dados de pesquisa

Dimensão	Categoria	Propriedade
Disposição para compartilhar dados	Financiamento público	Dado publicamente acessível
Fator estimulador à abertura de dados	Exigência editorial	Requisito para publicação
Fator inibidor à abertura de dados	Extremo zelo	Apego a dados
Obrigatoriedade pela universidade	Desafios técnicos	Falta de infraestrutura
Obrigatoriedade por agências de fomento	Diferenças disciplinares	Impossibilidade de generalizações
Obrigatoriedade por editores científicos	Favorável	Depósito obrigatório

Fonte: elaborado pela autora

4.3.4.2 Codificação aberta da entrevista com PQu2

PQu2 afirmou ter disposição para compartilhar os dados de suas pesquisas desde que:

- não haja mais perspectiva de geração de tecnologia;
- não haja negociação do projeto de pesquisa com alguma instituição privada interessada em explorá-lo;
- não haja temor de que dados possam ser copiados ou transferidos para pessoas não autorizadas;
- dados não apresentem aspecto inovador, podendo representar um ganho institucional em termos de divisas.

Ao falar sobre fatores que estimulam a abertura de dados de pesquisa, PQu2 mencionou que dados abertos podem dar mais visibilidade à pesquisa, e, conseqüentemente, atrair potenciais parcerias na sociedade para o desenvolvimento de estudos, principalmente sobre o ponto de vista de investimentos.

Quanto a fatores inibidores, PQu2 disse que a abertura de dados provenientes de estudos não concluídos pode ocasionar a perda do ineditismo de descobertas científicas. Conforme explicado por PQu2, há um desequilíbrio no processo de produção científica, uma vez que cientistas que dispõem de maior infraestrutura poderiam reutilizar dados e publicar resultados de pesquisas antes mesmo do pesquisador que os gerou.

Ao discorrer sobre a possibilidade de a universidade em que atua tornar obrigatório o depósito de dados de pesquisa, PQu2 afirmou que dados não poderiam ser disponibilizados, enquanto os estudos que os geraram não fossem concluídos. Ou seja, enquanto a pesquisa estivesse sendo desenvolvida, não se saberia qual o impacto científico que poderia ser alcançado com os dados produzidos. No contexto de agências de fomento, PQu2 disse que é favorável, desde que não se trate de dados confidenciais. Já no âmbito de editores científicos, PQu2 mencionou que a abertura de dados é uma questão consensual entre pesquisadores e os próprios editores.

Por fim, PQu2 destacou a necessidade da criação de diretrizes institucionais que demonstrem como a disponibilização de dados abertos de pesquisa deve ocorrer. Nesse sentido, percebe-se a ausência de uma instituição brasileira que aponte como um expoente de orientação para pesquisadores em relação aos dados que produzem, conforme explicou PQu2:

Então, eu acho que talvez, sei lá, se fosse o INPE [Instituto Nacional de Pesquisas Espaciais], fosse o Ministério da Ciência e da Tecnologia, fosse, eu não sei, teria que ter ABNT [Associação Brasileira de Normas Técnicas] ou um Inmetro [Instituto Nacional de Metrologia, Qualidade e Tecnologia], enfim, um órgão que fosse ser o responsável por dar o tom. Ele é quem vai dar o tom desse processo e vai contar com todos esses colaboradores satélites para chegar em uma normativa, para chegar em um conceito, em uma definição e em um modus operandi disso. Certo? Então, o que vem a ser, vamos definir divulgação aberta de dados de pesquisa, vamos ver a quem isso interessa, como isso pode ser efetivamente realizado, quem são os atores, como que eles vão ser auxiliados, como que eles vão ser, como que vai ser dado o suporte, isso vai ser feito, quem é quem vai obrigar, quem que não vai poder obrigar, em que condições.

O Quadro 59 apresenta as dimensões, categoria e propriedades identificadas na codificação aberta, de acordo com a entrevista de PQu2.

Quadro 59 – Codificação aberta da entrevista com PQu2: fatores que influenciam a abertura de dados de pesquisa

Dimensão	Categoria	Propriedade
Disposição para compartilhar dados	Divulgação condicionada	Divulgação de dados com restrições
Fator estimulador à abertura de dados	Visibilidade da pesquisa	Possibilidade de parcerias
Fator inibidor à abertura de dados	Perda de ineditismo	Desvantagem na publicação de dados
Obrigatoriedade pela universidade	Pesquisa em andamento	Resultados em desenvolvimento
Obrigatoriedade por agências de fomento	Favorável	Disponibilização de dados não confidenciais
Obrigatoriedade por editores científicos	Acordos	Negociação entre pesquisadores e editores
	Instituições responsáveis	Ausência de orientações

Fonte: elaborado pela autora

4.3.4.3 Codificação aberta da entrevista com PQu3

No que se refere à abertura de dados, PQu3 entende que já os publica no corpo de trabalhos científicos depositados em alguma plataforma ou na forma de patentes, descrevendo-os, principalmente, na seção de materiais e métodos. Além disso, comentou que há muitas incertezas no contexto da abertura de dados, no qual não se sabe ao certo quais são os reais interesses que cerca esse movimento. Sendo assim, PQu3 fez os seguintes questionamentos:

Primeiro, é o seguinte: a quem interessa a ciência aberta? É uma ideia de quem? De que país? De onde vem? É uma coisa de país de terceiro mundo ou é uma coisa de país de primeiro mundo?

Ao falar sobre fatores que estimulam a abertura de dados entre pesquisadores, PQu3 citou a natureza de dados que tratem de inovação. Outros tipos de dados que não possuem esse caráter, poderiam ser disponibilizados, principalmente se há desconfiança sobre como o trabalho científico foi realizado. No que se refere a fatores inibidores, PQu3 apontou o acesso não autorizado a dados, o que gera insegurança para o pesquisador, no sentido de outras pessoas poderem se apropriar indevidamente de dados.

Com relação à possibilidade de universidades, agências de fomento e editores científicos tornarem obrigatório o depósito de dados abertos, PQu3 disse que é contra tal medida, uma vez que a abertura de dados deveria ser uma concessão voluntária

do pesquisador. Além disso, como se trata de uma propriedade intelectual, PQu3 sugeriu que todo o processo de abertura de dados fosse amparado por uma legislação específica, pois sem isso, seria algo muito arriscado ao pesquisador.

O Quadro 60 apresenta as dimensões, categorias e propriedades identificadas na codificação aberta, com base na entrevista com PQu3.

Quadro 60 – Codificação aberta da entrevista com PQu3: fatores que influenciam a abertura de dados de pesquisa

Dimensão	Categoria	Propriedade
Disposição para compartilhar dados	Dados publicados	Publicação de dados em trabalhos científicos
	Incertezas	Interesses sobre a publicação de dados
Fator estimulador à abertura de dados	Natureza dos dados	Dados que não tratem de inovação tecnológica
Fator inibidor à abertura de dados	Acesso não autorizado	Insegurança sobre níveis de acesso a dados
Obrigatoriedade pela universidade		
Obrigatoriedade por agências de fomento	Concessão voluntária	Decisão do pesquisador
Obrigatoriedade por editores científicos		

Fonte: elaborado pela autora

4.3.4.4 Codificação aberta da entrevista com PQu4

PQu4 disse que tem disposição para compartilhar dados, entretanto, não consegue fazê-lo em razão dos altos custos para a publicação em acesso aberto. Nesse sentido, PQu4 apontou que a previsão do custeio da abertura de dados em editais de financiamento à pesquisa é um fator que estimularia a abertura de dados entre pesquisadores. Bem como o alto custo para a publicação em acesso aberto, associado à indisponibilidade de recursos é um fator que a inibe.

Ao discorrer sobre a possibilidade de a universidade em que atua tornar obrigatório o depósito de dados abertos, PQu4 explicou que as pesquisas que realiza são realizadas com o dinheiro público, logo, a sociedade deveria ter acesso ao conhecimento produzido, inclusive, aos dados. No contexto de agências de fomento, PQu4 concordou com a medida e afirmou que isso ampliaria a divulgação de pesquisas geradas no Brasil, de modo a incluir relatórios intermediários e avaliações

de projetos de pesquisa. Com relação aos editores científicos, PQu4 relatou não acreditar que revistas científicas deixem de cobrar pela publicação em acesso aberto, uma vez que essa não é custeada pelo Estado.

Mais uma vez, PQu4 enfatizou que a disponibilização de recursos é fundamental para o compartilhamento de dados, uma vez que não caberia ao pesquisador custear esse processo. Sobre isso, PQu4 reiterou:

Alguém tinha que pagar e o custo é muito caro. Então, deveria ter um financiamento da própria instituição ou do ministério da educação para poder pagar esse acesso aberto, né? [...] e para pesquisar, eu preciso de recursos, eu preciso de um planejamento que vem do governo para poder disponibilizar, porque com 30 mil reais eu não faço muita coisa não! Não compro muitos reagentes, são muito caros, eu não compro um equipamento para colocar dentro do laboratório, é 10 mil reais! Então, já vai um terço aí do valor do financiamento, né? Eu tenho aqui equipamentos de amostragem de ar, que eles custam em torno de 12 mil reais. Então, se eu montar um projeto e se eu não tiver já esse equipamento hoje, eu não consigo montar um projeto para poder trabalhar, por causa do recurso. O recurso é primordial aí, dentro da distribuição desse recurso, para que a gente possa prosseguir fazendo pesquisa.

O Quadro 61 apresenta as dimensões, categorias e propriedades identificadas na codificação aberta, de acordo com a entrevista com PQu4.

Quadro 61 – Codificação aberta da entrevista com PQu4: fatores que influenciam a abertura de dados de pesquisa

Dimensão	Categoria	Propriedade
Disposição para compartilhar dados	Custos	Publicação mediante pagamento
Fator estimulador à abertura de dados	Editais	Custeio da abertura de dados
Fator inibidor à abertura de dados	Custos	Indisponibilidade de recursos
Obrigatoriedade pela universidade	Dinheiro público	Acesso público a pesquisas financiadas com dinheiro público
Obrigatoriedade por agências de fomento	Ampla divulgação	Disponibilização pública de dados de pesquisa
Obrigatoriedade por editores científicos	Publicação mediante pagamento	Pagamento de autores para a publicação

Fonte: elaborado pela autora

4.3.4.5 Codificação aberta da entrevista com PQu5

PQu5 afirmou que tem disponibilidade para compartilhar dados abertos de pesquisa, desde que haja critérios que estabeleçam algumas restrições a dados de caráter tecnológico. Tal medida é imprescindível para mitigar riscos de eventuais

concorrentes desenvolverem mais rapidamente produtos, em razão do acesso antecipado a dados. Nessa direção, PQu5 explicou que o ritmo de produção científica em universidades é diferente daquele encontrado no contexto corporativo:

Desde que os critérios para esse tipo de processo sejam muito bem estabelecidos, exista restrições para alguns tipos de pesquisa de caráter tecnológico, sim! Onde eu quero chegar com isso? Se eu estou desenvolvendo um produto, onde eu vou ter certeza que um eventual concorrente não vai fazer isso mais rápido do que eu? Como eu te disse, eu estou desenvolvendo dentro do âmbito da universidade e a velocidade é muito diferente dentro do âmbito corporativo. Então, eu acho que sim, se condicionada aos critérios.

Quanto a fatores que estimulariam a abertura dos dados de pesquisa, PQu5 citou a cooperação científica. Trata-se da situação na qual pesquisadores liberam seus dados, mas também têm acesso a dados de outros. Quanto a fatores inibidores a essa prática, PQu5 mencionou a concorrência, seja por questões de mercado, seja por aprovação de projeto ou de recursos. Ao relacionar cooperação com concorrência, PQu5 fez o seguinte comentário:

Então, como é que a gente vai balancear essas duas questões? Cooperação e concorrência, até que ponto quando eu estou liberando os meus dados eu estou ficando mais fraco do ponto de vista de concorrência? Ou o contrário? Então, como eu te disse, tudo vai perpassar pelos critérios. Eu posso tornar a liberação de dados tanto interessante do ponto de vista de concorrência, como do ponto de vista de colaboração. Ou seja, eu posso tornar a pessoa que libera seus dados concorrente mais forte do que aquele que não libera, mas isso para mim ainda depende de como é que os critérios ainda vão ser estabelecidos.

Ao dar a sua opinião sobre a possibilidade de a universidade em que atua tornar obrigatório o depósito de dados abertos, PQu5 disse que está de acordo com tal medida, na medida em que se estabeleça um prazo razoável para a implementação de novas rotinas para esse propósito, e que haja plataformas específicas que viabilizem a abertura de dados. Entretanto, destacou que a palavra “obrigatoriedade” geralmente é mal recebida pelo ser humano, sendo muito provável que pesquisadores venham a reclamar sobre imposições nesse sentido.

No que se refere à possível obrigatoriedade no contexto de agências de fomento, PQu5 enxerga um caminho menos dificultoso, uma vez que o pesquisador já saberia que se não disponibilizasse abertamente seus dados, não obteria o recurso para o financiamento da pesquisa. Dessa forma, caberia ao pesquisador a decisão de aceitar, ou não, os termos estabelecidos pelas agências em relação aos dados. Já em relação aos editores científicos, PQu5 relatou que esse é um modelo

relacionado à comprovação de dados mediante a verificação da fidedignidade dos mesmos e à transparência nos processos de pesquisas.

O Quadro 62 apresenta as dimensões, categorias e propriedades identificadas na codificação aberta, conforme a entrevista com PQu5.

Quadro 62 – Codificação aberta da entrevista com PQu5: fatores que influenciam a abertura de dados de pesquisa

Dimensão	Categoria	Propriedade
Disposição para compartilhar dados	Dados de caráter tecnológico	Restrições para mitigar riscos de exposição a concorrentes
Fator estimulador à abertura de dados	Cooperação científica	Possibilidade de acessar dados de outros pesquisadores
Fator inibidor à abertura de dados	Concorrência	Questões de mercado, aprovação de projetos ou de recursos
Obrigatoriedade pela universidade	Obrigatoriedade	A palavra “obrigatório” pode implicar em receio de pesquisadores quanto à abertura de dados
Obrigatoriedade por agências de fomento	Decisão do pesquisador	Financiamento à pesquisa condicionado à abertura de dados
Obrigatoriedade por editores científicos	Comprovação de dados	Verificação da fidedignidade de dados e transparência nos processos de pesquisas

Fonte: elaborado pela autora

4.3.4.6 Codificação aberta da entrevista com PQu6

PQu6 disse que tem disposição para compartilhar dados abertos de pesquisa, e que só não publica mais em acesso aberto, devido aos altos custos relacionados à publicação. Dentre os fatores que estimulariam a abertura de dados por parte de pesquisadores, PQu6 mencionou o custeio institucional da publicação de dados abertos, uma vez que não caberia ao pesquisador arcar com nenhum encargo financeiro dessa natureza. Logo, o principal fator inibidor apontado por PQu6 foi a cobrança de taxas para a publicação de conteúdos em acesso aberto. Sobre isso, PQu6 explicou:

Eu acho que tem que ter estímulo à pesquisa de uma forma geral no Brasil e isso passaria por divulgação dos dados, tudo em open access. Só que não a cargo do pesquisador para pagar a taxa, [mas sim] a cargo da instituição ou da agência de fomento. Então, eu acho que aí acaba com essa história de assinatura, todo mundo tem acesso a tudo. Porque na verdade, a gente que faz pesquisa usando verba pública, tudo bem que alguém que não é da área não vai entender, mas eu acho que tem que ser divulgado, tem que mostrar.

Ao se referir à possibilidade de a universidade em que atua e de agências de fomento tornarem obrigatório o depósito de dados abertos de pesquisa, PQu6 argumentou que por se tratar de pesquisas financiadas por dinheiro público, os dados obtidos também deveriam ser publicamente acessíveis. Nesse sentido, PQu6 comentou que a preocupação sobre a disponibilidade de dados deveria partir do próprio pesquisador.

PQu6 também concorda com uma possível obrigatoriedade de depósito de dados abertos de pesquisa no contexto de editores científicos, principalmente com a finalidade de verificar a idoneidade de estudos. PQu6 disse que há muitas revistas predatórias que aceitam qualquer trabalho, sem o devido processo de revisão por pares, somente para o aumento no número de suas publicações. Assim, tal obrigatoriedade seria um recurso útil para a comprovação de dados obtidos e verificação da qualidade desses dados.

Por fim, PQu6 destacou que, ao longo de sua carreira acadêmica, percebeu um avanço científico impulsionado por novas formas de comunicação do conhecimento, podendo-se destacar a publicação de dados abertos de pesquisa. De acordo com PQu6, apesar de as universidades ainda necessitarem avançar em relação à discussão e promoção do acesso aberto, tem sido cada vez mais frequente a disponibilização de dados por meio de repositórios, bancos de dados e outras plataformas digitais. Por essa razão, PQu6 reiterou que tal temática deve ser melhor discutida no Brasil para ampliar novas possibilidades de acesso e uso do conhecimento científico.

O Quadro 63 apresenta as dimensões, categorias e propriedades identificadas na codificação aberta, conforme o relato de PQu6.

Quadro 63 – Codificação aberta da entrevista com PQu6: fatores que influenciam a abertura de dados de pesquisa

Dimensão	Categoria	Propriedade
Disposição para compartilhar dados	Custos	Publicação mediante pagamento
Fator estimulador à abertura de dados	Custeio institucional	Custeio da abertura de dados
Fator inibidor à abertura de dados	Custos	Taxas para a publicação
Obrigatoriedade pela universidade		
Obrigatoriedade por agências de fomento	Dinheiro público	Acesso aberto a dados de pesquisas financiadas com dinheiro público
Obrigatoriedade por editores científicos	Idoneidade da pesquisa	Comprovação da veracidade de dados de pesquisa

Fonte: elaborado pela autora

4.3.4.7 Codificação aberta da entrevista com PQu7

PQu7 afirmou que está disposto a compartilhar dados das pesquisas que realiza, mas somente após essas forem concluídas. Isso, para evitar a perda do ineditismo de publicações em um contexto altamente competitivo, já que dados podem originar novas pesquisas. Nessa direção, PQu7 explicou:

Eu acho que só quando tiver finalizado, porque algumas pesquisas dão origem a outras pesquisas, e daí, às vezes, fica complicado, porque há muita concorrência no meio, e daí, quem faz o produto final é quem vai ser o pai da criação. Eu acho que tem que ter uma parcimônia, algumas coisas a gente pode liberar, e outras, eu acho que tem que esperar até chegar ao final.

Para PQu7, o risco de furto de ideias é um fator inibidor para a abertura de dados de pesquisa, ou seja, o receio de que ocorra a apropriação indevida de dados por outros pesquisadores. Com relação a fatores que poderiam estimular tal prática, PQu7 destacou a divulgação científica, que proporcionaria maior interação entre o conhecimento produzido em universidades e sua aplicação no contexto social. Nesse sentido, PQu7 explicou que a abertura de dados é uma maneira de mostrar que cientistas estão desenvolvendo novos produtos a serviço da sociedade, por meio do dinheiro público:

A gente tem que mostrar para a sociedade que a gente está fazendo alguma coisa. Às vezes, eles não entendem: ah, vai sair um biocombustível! Vai sair um eletrodo novo! Vai sair um catalisador que vai acelerar a reação x! Então, a sociedade como um todo não entende que para chegar naquele produto final, tu tens vários passos a percorrer. Mas eu acho que o pesquisador tem obrigação de mostrar que ele está fazendo um bem para a sociedade, que ele não está simplesmente ganhando um dinheiro público em vão.

Ao comentar sobre a possibilidade de a universidade tornar obrigatório o depósito de dados abertos de pesquisa, PQu7 destacou a necessidade de divulgação de pesquisas realizadas na própria instituição, uma vez que muitos dados encontram-se inacessíveis ou apresentam muitas restrições para serem consultados. Para PQu7, geralmente, os dados produzidos em sua universidade ficam circunscritos ao grupo de pesquisa que integra, não havendo nenhum banco de dados ou plataforma destinada à armazenagem e disponibilização de dados de pesquisa.

No que se refere a tal obrigação no contexto de agências de fomento, PQu7 afirmou que dados obtidos mediante financiamento público deveriam ser abertos, até porque, trata-se de uma prestação de contas à sociedade sobre os recursos investidos nas pesquisas realizadas. Já em relação à possível obrigatoriedade estabelecida por editores científicos, PQu7 disse que ainda há muitas limitações editoriais que inviabilizam uma abertura de dados de modo eficaz, principalmente relacionadas a aspectos tecnológicos.

O Quadro 64 apresenta as dimensões, categorias e propriedades identificadas na codificação aberta, conforme a entrevista com PQu7.

Quadro 64 – Codificação aberta da entrevista com PQu7: fatores que influenciam a abertura de dados de pesquisa

Dimensão	Categoria	Propriedade
Disposição para compartilhar dados	Ineditismo da publicação	Criação de produtos
Fator estimulador à abertura de dados	Divulgação científica	Interação entre universidade e sociedade
Fator inibidor à abertura de dados	Furto de ideias	Apropriação indevida de ideias
Obrigatoriedade pela universidade	Dados de acesso restrito	Necessidade de divulgação de pesquisas
Obrigatoriedade por agências de fomento	Dinheiro público	Abertura de dados produzidos mediante financiamento público
Obrigatoriedade por editores científicos	Inviabilidade	Limitações editoriais

Fonte: elaborado pela autora

4.3.4.8 Codificação aberta da entrevista com PQu8

PQu8 disse que estava disposto a compartilhar dados de pesquisa: “com toda a certeza do mundo, mas com toda mesma!” Além disso, reiterou que o acesso a

dados poderia ser ponte para o desenvolvimento de novas pesquisas direcionadas, até mesmo, para contextos diferentes do estudo de origem.

PQu8 elencou os seguintes fatores que incentivariam a abertura de dados pelo pesquisador:

- pontuação na progressão da carreira;
- exigência da abertura de dados para o recebimento de financiamento concedido por agências de fomento;
- exigência da abertura de dados para o recebimento de bolsa produtividade paga pelo CNPq.

Semelhantemente, PQu8 mencionou os seguintes fatores que inibiriam a abertura de dados pelo pesquisador:

- falta de incentivo para o compartilhamento de dados abertos;
- vaidade do pesquisador, por não admitir que outro reutilize dados originais antes dos mesmos serem publicados;
- receio de ideias serem roubadas.

Por considerar que o receio do furto de ideias é o fator que mais inibiria a abertura de dados, PQu8 explicou:

A pessoa trabalha com isso e tem outras pessoas que roubam ideia mesmo, isso acontece. Eu acho que isso aí é um dos principais fatores que inibem os pesquisadores a abrirem os dados [...] os pesquisadores têm essa relutância mesmo, por sei lá, talvez seja competitividade mesmo, medo de competição, esse tipo de coisa.

Ao falar sobre a possibilidade de a universidade em que atua torne obrigatório o depósito de dados abertos de pesquisa, PQu8 é a favor de tal medida, desde que os dados não identifiquem pessoas ou não necessitem de sigilo por algum motivo relacionado à pesquisa. PQu8 também discorreu sobre a gestão de dados abertos de pesquisa e sugeriu que houvesse um repositório nacional de dados abertos que abrangesse os dados de diversas universidades do país, de modo sistematizado.

Com relação ao contexto de agências de fomento, PQu8 destacou que é totalmente a favor da obrigatoriedade do depósito de dados, por se tratar de uma prestação de contas do cientista à sociedade. Assim também, PQu8 é favorável à obrigatoriedade estabelecida por editores científicos, principalmente para verificar a

fidedignidade de dados e atestar sua contribuição científica. Para PQu8, isso poderia diminuir a quantidade de resultados de pesquisas equivocadas publicados atualmente.

O Quadro 65 apresenta as dimensões, categorias e propriedades identificadas na codificação aberta, conforme a entrevista com PQu8.

Quadro 65 – Codificação aberta da entrevista com PQu8: fatores que influenciam a abertura de dados de pesquisa

Dimensão	Categoria	Propriedade
Disposição para compartilhar dados	Proposição de novos estudos	Utilização de dados em novos estudos
	Sistema de recompensas	Pontuação na progressão da carreira
Fator estimulador à abertura de dados	Obrigatoriedade	Exigência da abertura de dados para o recebimento de financiamento e concessão de bolsa-produtividade
	Falta de incentivo ao pesquisador	Desestímulo no compartilhamento de dados
Fator inibidor à abertura de dados	Vaidade do pesquisador	Não autorização do reuso de dados de pesquisa
	Furto de ideias	Apropriação indevida de ideias
	Dados com restrições éticas	Preservação da identidade de participantes da pesquisa
Obrigatoriedade pela Universidade	Repositório nacional de dados abertos	Sistema nacional que compila dados das universidades brasileiras
	Prestação de contas	Prestação de contas do cientista à sociedade
Obrigatoriedade por agências de fomento	Comprovação de dados	Verificação da fidedignidade de dados

Fonte: elaborado pela autora

4.3.4.9 Codificação aberta da entrevista com PQu9

PQu9 afirmou que tem disposição para compartilhar os dados que produz, desde que aqueles que venham a utilizá-los assinem um termo com condições específicas de uso e que atendam à exigência de citação do laboratório responsável pela geração dos dados.

Dentre os fatores que estimulariam a abertura de dados de pesquisa na área, PQu9 citou a criação de um programa liderado por uma instituição de representação mundial que viabilizasse um consórcio de dados entre comunidades científicas de

diferentes países, e que a mesma se responsabilizasse com os custos envolvidos. Segundo PQu9, as áreas que integram as Ciências Exatas ainda trabalham de forma seccionada quanto ao compartilhamento de dados abertos.

Já quanto aos fatores que poderiam inibir a abertura de dados de pesquisa na área, PQu9 mencionou a perda do ineditismo, ocasionada por uma perda no valor científico de dados, uma vez que esses já teriam sido publicados abertamente. Sob tal ponto de vista, algumas revistas científicas ficariam desinteressadas em publicar dados que não fossem inéditos, conforme apontou PQu9:

Vou dar um exemplo: a American Chemical Society, Royal Society of Chemistry, Sociedade Brasileira de Química têm suas editoras e têm as suas revistas que são conceituadas e se eu abrir o meu dado hoje, eu não posso mais submeter para essas revistas porque ele é considerado um dado aberto e ele não tem mais valor de publicação e isso gera uma cascata de problemas.

Ao falar sobre uma possível obrigatoriedade de depósito de dados abertos instituída por universidades e agências de fomento, PQu9 reiterou que tal obrigação contribui para a perda do ineditismo na ciência, causando prejuízos a pesquisadores. Quanto a essa possível obrigatoriedade no contexto de editores científicos, PQu9 disse que não há muito o que se fazer no sentido da abertura de dados, uma vez que o autor cede os direitos autorais às revistas, que passam a ser detentoras de todo o conteúdo produzido.

Para mitigar os riscos relacionados à perda de ineditismo devido à publicação antecipada de dados de pesquisa, PQu9 destacou a importância da criação de mecanismos de proteção a dados. Nessa direção, sugeriu que houvesse um período de embargo que antecederesse o acesso a dados abertos de pesquisa, semelhantemente como ocorre no patenteamento de registros de processos e produtos.

O Quadro 66 apresenta as dimensões, categorias e propriedades identificadas na codificação aberta, conforme a entrevista com PQu9.

Quadro 66 – Codificação aberta da entrevista com PQu9: fatores que influenciam a abertura de dados de pesquisa

Dimensão	Categoria	Propriedade
Disposição para compartilhar dados	Termo de uso de dados	Disponibilização de dados mediante cumprimento de termos
Fator estimulador à abertura de dados	Programas internacionais	Consórcio de dados
Fator inibidor à abertura de dados	Perda de ineditismo	Perda no valor de publicação
Obrigatoriedade pela universidade	Perda de ineditismo	Prejuízos na publicação em razão da perda de ineditismo
Obrigatoriedade por agências de fomento		
Obrigatoriedade por editores científicos	Direitos autorais	Cessão de direitos a revistas
	Período de embargo	Prazo que antecede o acesso a dados abertos de pesquisa

Fonte: elaborado pela autora

4.3.4.10 Codificação aberta da entrevista com PQu10

PQu10 disse que está disposto a compartilhar dados abertos de pesquisa, desde que não apresentem caráter de inovação tecnológica. Isso porque dados dessa natureza geralmente necessitam ser resguardados em sigilo devido a pactos contratuais ou intenção de parcerias com instituições privadas, ou por serem insumos para o registro de patentes. Nesse sentido, PQu10 explicou:

Dependendo, sim! Agora a gente tem que resguardar quando a gente está fazendo alguma coisa, por exemplo, sigilosa. Agora acabamos de fazer uma patente com a França, está sob sigilo, eu não posso nem divulgar esse dado agora porque tem um acordo sobre isso. Estamos buscando um parceiro comercial, como eu vou divulgar esse dado? Então depende do dado que você está falando, se for um dado mais corrente de interesse da sociedade, eu acho que os caras têm obrigação de divulgar e eu não me furtaria a isso.

Do ponto de vista de PQu10, a disponibilidade de recursos financeiros, tecnológicos e de pessoal é um dos fatores que poderiam estimular a abertura de dados. Nesse sentido, PQu10 destacou a necessidade da criação e manutenção de plataformas de acesso aberto, além de equipes capacitadas para viabilizar sistematicamente tal abertura. De outro ângulo, PQu10 disse que a natureza do dado é um aspecto que poderia inibir essa prática. Desse modo, dados direcionados à sociedade, como aqueles que abrangem a aferição da qualidade da água e do ar, deveriam ser abertos. Todavia, dados estratégicos que poderiam inserir o Brasil em

uma vanguarda tecnológica teriam que aguardar por um período até que pudessem ser publicados.

Ao falar sobre a possibilidade de a universidade em que atua tornar obrigatório o depósito de dados abertos, PQu10 disse que é favorável a essa medida. Entretanto, mais uma vez, destacou que há dados que necessitam ser resguardados até a sua publicação, por serem insumos de potencial registro de patentes. Assim, PQu10 fez o seguinte comentário:

Qualquer um que escolhe trabalhar em uma universidade, ele escolhe doar, doar conhecimento, doar informação, formar pessoas, etc. E eu não sou diferente disso, não! Então, eu acho que você deve ter a obrigação de publicar os teus dados, mas tem que também resguardar o momento de fazer aquilo, e isso cabe a cada cientista, a cada pesquisador. Você não pode: 'a universidade vai lá no departamento e me dá esse teu resultado cromatográfico', eu falei 'calma, isso pode virar uma patente para você daqui há um ano, então espera que eu estou fazendo, estou trabalhando'.

Quanto ao seu posicionamento em relação à obrigatoriedade de depósito de dados abertos estabelecida por agências de fomento, PQu10 também afirmou ser favorável a tal medida. Porém, destacou o risco ético de outros pesquisadores furtarem ideias e as publicarem como sendo de sua autoria. Nessa direção, PQu10 explicou:

Então, tem que tomar cuidado com isso! Ninguém pode chegar em um congresso e vai falar: 'olha, eu descobri isso, faz assim, assim assado, que você chega em tal lugar!' Você não pode fazer isso com risco de ser copiado, não seria legal. A gente já tem problema de citação, de gente que faz a mesma coisa que você e não cita o teu trabalho, quer dizer, isso é um dado, o cara teve acesso ao teu dado, então, o problema de ética na ciência ele existe e existe no Brasil e fora do Brasil.

Por fim, com relação a tal obrigatoriedade no contexto de editores científicos, PQu10 disse que publica os dados de pesquisa na seção de material suplementar em quase toda publicação que realiza em periódicos científicos.

O Quadro 67 apresenta as dimensões, categorias, subcategorias e propriedades identificadas na codificação aberta, conforme a entrevista com PQu10.

Quadro 67 – Codificação aberta da entrevista com PQu10: fatores que influenciam a abertura de dados de pesquisa

Dimensão	Categoria	Propriedade
Disposição para compartilhar dados	Dados de caráter tecnológico	Pactos contratuais com instituições privadas
Fator estimulador à abertura de dados	Recursos	Disponibilidade de recursos financeiros, tecnológicos e de pessoal
Fator inibidor à abertura de dados	Natureza dos dados	Dados com caráter de inovação tecnológica
Obrigatoriedade pela universidade	Dados com caráter de inovação tecnológica	Dados que necessitam ser resguardados até a sua publicação
Obrigatoriedade por agências de fomento	Riscos éticos	Furto de ideias
Obrigatoriedade por editores científicos	Material suplementar	Publicação de dados juntamente com artigos

Fonte: elaborado pela autora

4.3.4.11 Síntese da codificação aberta

O processo de codificação aberta nas entrevistas realizadas resultou em uma descrição inicial dos fatores que influenciam a abertura de dados de pesquisa na área de Química. Sendo assim, foram identificadas 59 categorias abertas, distribuídas em seis dimensões. Essas categorias passaram por processo de codificação axial, convertendo-se em outras quatro categorias axiais, que, por sua vez, se desdobraram em dez subcategorias, as quais podem ser observadas no Quadro 68.

Quadro 68 – Síntese da codificação aberta – Química: fatores que influenciam a abertura de dados de pesquisa

Dimensão	Categoria aberta	Categoria axial	Subcategoria axial
Disposição para compartilhar dados	Financiamento público	Fatores estimuladores à abertura de dados	Aprimoramento de estudos na área
	Divulgação condicionada		Liderança institucional
	Dados publicados		Aumento na visibilidade de pesquisas
	Incertezas		Recursos destinados à abertura de dados
	Custos	Fatores inibidores à abertura de dados	Dados de natureza tecnológica
	Dados de caráter tecnológico		Riscos éticos
	Ineditismo da publicação		Custos da disponibilização de dados abertos de pesquisa
	Proposição de novos estudos	Favorável à obrigatoriedade	Pesquisas fomentadas com dinheiro público
	Termo de uso de dados		Comprovação de dados por editores científicos
	Fator estimulador à abertura de dados	Exigência editorial	Desfavorável à obrigatoriedade
Visibilidade da pesquisa			
Natureza dos dados			
Editais			
Cooperação científica			
Custeio institucional			
Divulgação científica			
Sistema de recompensas			
Obrigatoriedade			
Programas Internacionais			
Recursos			
Fator inibidor à abertura de dados	Extremo zelo		
	Perda de ineditismo		
	Acesso não autorizado		
	Custos		
	Concorrência		
	Furto de ideias		
	Falta de incentivo ao pesquisador		
	Vaidade do pesquisador		
Natureza dos dados			

Obrigatoriedade pela universidade	Desafios técnicos
	Pesquisa em andamento
	Concessão voluntária
	Dinheiro público
	Obrigatoriedade
	Dados de acesso restrito
	Dados com restrições éticas
	Repositório nacional de dados abertos
	Perda de ineditismo
	Dados com caráter de inovação tecnológica
Obrigatoriedade por agências de fomento	Diferenças disciplinares
	Favorável
	Concessão voluntária
	Ampla divulgação
	Decisão do pesquisador
	Dinheiro público
	Prestação de contas
	Perda de ineditismo
Riscos éticos	
Obrigatoriedade por editores científicos	Favorável
	Acordos
	Concessão voluntária
	Publicação mediante pagamento
	Comprovação de dados
	Idoneidade da pesquisa
	Inviabilidade
	Direitos autorais
	Período de embargo
	Material suplementar
Instituições responsáveis	

Fonte: elaborado pela autora

4.3.5 Codificação axial – Química

A presente seção relacionará as categorias e subcategorias identificadas no estudo, a partir do processo de codificação axial realizado.

4.3.5.1 Categoria “Fatores estimuladores à abertura de dados de pesquisa”

A categoria “Fatores estimuladores à abertura de dados de pesquisa” contempla os aspectos identificados no estudo que cooperam para o incentivo da abertura dos mesmos. Em suma, tal categoria abrange o conjunto de forças que atuam sobre pesquisadores, encorajando-os, de maneira a influenciar o seu comportamento em direção à abertura de dados.

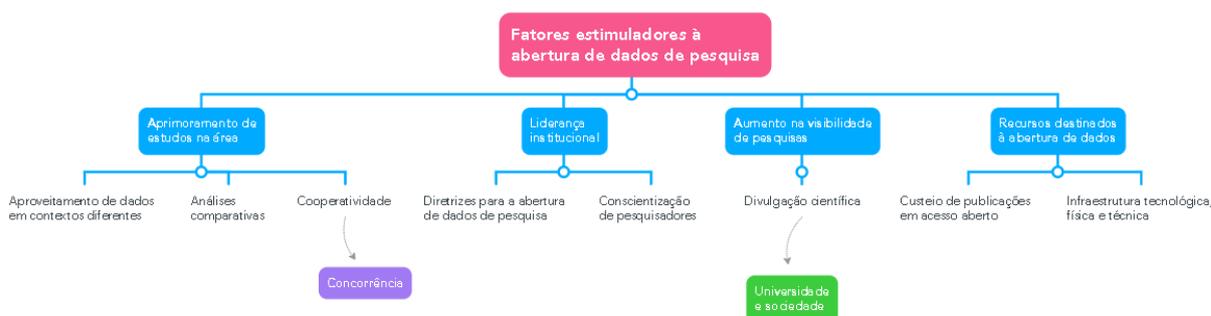
Como pode ser visto na Figura 46, esta categoria se desenvolveu nos termos de quatro subcategorias, sendo elas: aprimoramento de estudos na área; liderança institucional; aumento na visibilidade de pesquisas; 4) recursos destinados à abertura de dados. A primeira diz respeito ao crescimento teórico e prático da área devido o aproveitamento de dados em contextos diferentes daqueles que os gerou, à possibilidade de análises comparativas e à cooperatividade entre pesquisadores, apesar da concorrência entre esses, ora pela demanda de mercado, ora pelo recebimento de recursos.

A segunda subcategoria refere-se à necessidade apontada por pesquisadores da ascensão de uma liderança institucional que possa estabelecer diretrizes para a abertura de dados de pesquisa no país. Para tanto, foi sugerida a criação de um consórcio de dados o qual proporcionaria o intercâmbio de dados originários de diferentes países. Ademais, tal liderança poderia promover campanhas de conscientização a respeito da importância do compartilhamento de dados abertos de pesquisa.

Por seu turno, a terceira subcategoria trata da possibilidade de aumento na visibilidade de pesquisas. Nesse contexto, a divulgação científica proporcionaria maior interação entre o conhecimento produzido em universidades e sua aplicação na sociedade. Por fim, a última subcategoria refere-se à importância da disponibilidade de recursos destinados à abertura de dados. Está relacionada ao custeio de publicações em acesso

aberto e à disposição de infraestrutura tecnológica, física e técnica necessária aos processos inerentes à abertura de dados.

Figura 46 – Categoria Fatores estimuladores à abertura de dados de pesquisa - Química



Fonte: elaborado pela autora

- Subcategoria “Aprimoramento de estudos na área”

Com base nas informações obtidas no estudo, pode-se observar que alguns entrevistados, principalmente PQu5 e PQu8, afirmaram que o acesso a dados abertos de pesquisa viabiliza o aprimoramento de estudos no campo da Química. Logo, a possibilidade do crescimento da área em nível teórico e prático foi considerado um fator de estímulo para a abertura de dados de pesquisa.

De acordo com PQu8, dados criados para fins de pesquisas específicas poderiam ser utilizados em contextos diferentes dos estudos de origem. Nesse sentido, PQu9 ilustrou a situação na qual dados coletados por meio de amostras sanguíneas para um estudo de diagnóstico de acidente vascular encefálico poderiam ser reinvestigados, sendo úteis, até mesmo, para o diagnóstico de outras doenças:

Por exemplo, a gente fez um trabalho visando se a gente conseguia fazer diagnósticos de acidente vascular encefálico, beleza, coletamos sangue para isso. Mas se outra pessoa pode pegar esses dados e reinvestigar esses dados e vai que ela descobre que ela poderia, que ela consegue diagnosticar outra doença? É totalmente válido. Sou absolutamente aberto mesmo.

De modo semelhante, PQu5 afirmou que é muito interessante o intercâmbio de dados entre pesquisadores, principalmente para a realização de comparações entre resultados obtidos. É nesse contexto que o sentido de “cooperatividade” entre cientistas se fortalece, e, de modo geral, a área da Química cresce em termos de pesquisas sendo executadas em um mesmo período. Trata-se de um princípio

colaborativo baseado na disponibilização de dados originais e no acesso a dados produzidos por outros pesquisadores, apesar da concorrência entre esses, seja pela demanda de mercado, seja pela aprovação de projetos, ou pela liberação de recursos.

À falar sobre isso, PQu5 fez as seguintes considerações:

Eu acho que o que vai estimular é: você liberar os seus e ter acesso aos dos outros. Então, é muito importante eu ter acesso aos dos outros para comparar com os meus. Mas, por outro lado, também na hora que estou abrindo os meus, eu posso estar motivando, dando informações muito relevantes a outros grupos de pesquisa. Eu acho que a gente tem dois conceitos que podem ser considerados: ciência aberta e cooperatividade. A cooperatividade é muito salutar, então, dentro do princípio de cooperação, eu ter acesso aos dados e eu liberar os meus dados. Só que a gente também tem que lembrar que existe o conceito, infelizmente, de concorrência.

- Subcategoria “Liderança institucional”

Segundo PQu3, a abertura de dados gerou um contexto de incertezas, no qual ainda se desconhecem os verdadeiros interesses que a cercam. Sendo assim, alguns questionamentos relacionados à origem do fenômeno da Ciência aberta e ao nível de abertura de dados em diferentes países permeiam o contexto de pesquisadores, o que sugere a necessidade de uma liderança institucional que norteie todo esse processo no país.

Nessa direção, PQu9 sugeriu a criação de um programa encabeçado por uma instituição de representação mundial que propusesse um consórcio de dados entre comunidades de pesquisadores de diferentes países, e que fosse a entidade responsável pelos custos inerentes aos processos de abertura de dados. Por isso, PQu9 exemplificou, em sua fala, a existência de programas internacionais que realizam consórcios para que pesquisadores possam realizar pesquisas a baixo custo. Em se tratando do contexto de dados de pesquisa, tal medida seria interessante juntamente com a criação de fundos internacionais capazes de manter um sistema público de disponibilização de dados, conforme explicou PQu9:

Eu acredito que o ideal seria a gente ter programas internacionais como tem os telescópios internacionais para observação do universo, eles têm programas internacionais aonde os países fazem um consórcio e os pesquisadores podem usar quase sem custo. Mas o custo vem de uma base internacional, digamos assim, é comum tu ter organização das nações unidas, tu ter uma base de apoio [...] talvez uma empresa privada vai pagar, mas a comunidade científica, em geral, ela não paga para fazer o experimento. Então, a constituição de fundos de investimento, ou fundos internacionais que pudessem manter esse sistema fornecendo dado público, seria bem interessante que fosse possível, uma organização única, porque para a ciência, hoje, é muito seccionada.

Semelhantemente, PQu2 mencionou a importância do estabelecimento de diretrizes institucionais a nível nacional que orientem como o pesquisador deve proceder em relação à disponibilização de dados abertos. Na perspectiva de PQu2, essas diretrizes poderiam ser definidas por instituições tais como o Instituto Nacional de Pesquisas Espaciais (INPE), o Ministério da Ciência e da Tecnologia, a Associação Brasileira de Normas Técnicas (ABNT) ou o Instituto Nacional de Metrologia, Qualidade e Tecnologia (Inmetro).

Conforme destacou PQu1, diretrizes voltadas à abertura de dados são relevantes para pacificar diferenças disciplinares relacionadas aos diversos contextos de produção nos quais dados são gerados. De outro modo, tais diretrizes também são relevantes para o propósito de conscientizar pesquisadores a respeito da importância do compartilhamento de dados abertos. De acordo com PQu8, falta incentivo ao pesquisador em relação à disponibilização dos dados que produz, bem como existe uma certa vaidade, na qual muitos cientistas não admitem que outros reutilizem dados originais antes desses serem publicados, sendo motivados, principalmente, pelo receio de furto de ideias. Sobre isso, PQu8 explicou:

Quando você publica os dados de uma molécula, beleza, vira público. Mas você publicar os seus dados originais para que outra pessoa possa reinvestigar, aí o pesquisador é bem relutante para fazer isso, pela vaidade mesmo, viu? Sei lá, eles acham que, por exemplo, os pesquisadores acreditam que o outro pesquisador vá roubar a ideia. De fato, isso acontece com bastante frequência, até aqui mesmo na universidade, eu vejo isso acontecer. Sei lá, a pessoa trabalha com isso e tem outras pessoas que roubam a ideia mesmo.

Por essa razão, PQu8 sugeriu a criação de um sistema de pontuação na progressão da carreira do pesquisador, no qual aquele que disponibilizasse dados abertos seria bonificado nesse sistema. Consequentemente, tal pesquisador teria mais oportunidades no que se refere à concessão de bolsas de pesquisa, entre outros benefícios. Sendo assim, PQu8 comentou:

O que incentivaria seria, eu acho, uma pontuação na progressão da carreira, incentivos na progressão da carreira, cobranças por parte das agências de fomento a pesquisas mesmo, colocar lá uma cláusula 'oh, vou deixar os seus dados abertos, senão você não vai ser financiado'. Outro exemplo aqui também que eu acho que incentiva os pesquisadores, eles têm uma bolsa de produtividade, que é paga pelo Cnpq, acho que também poderia haver uma cláusula lá, se você colocar os seus dados abertos você vai fazer jus a essa bolsa de oportunidade, e do contrário, não. Ou seja, algumas das formas de incentivo.

- Subcategoria “Aumento na visibilidade de pesquisas”

Com base nos relatos obtidos no estudo, pode-se observar que o aumento na visibilidade de pesquisas foi um dos fatores que estimulam a abertura de dados de pesquisa apontado pelos entrevistados. De acordo com PQu2, dados abertos proporcionam visibilidade à instituição em que atua e ao seu laboratório, de modo a demonstrar à sociedade os trabalhos que estão sendo realizados no ambiente acadêmico.

Nessa direção, PQu2 lembrou uma situação na qual um de seus projetos de pesquisa de enfrentamento à Covid foi noticiado na mídia, e a partir de então, recebeu uma doação de recursos para dar suporte à execução da pesquisa. Sobre isso, PQu2 explicou:

O que me estimula é dar visibilidade à minha instituição. É dar visibilidade ao meu laboratório de pesquisa, quanto ao que nós estamos fazendo, e que isso sirva não só de motivo para que a sociedade, de maneira geral, saiba e valorize o que nós estamos fazendo, como que também ocorra o despertar de potenciais parceiros da sociedade para poder contribuir para a ciência, principalmente, sobre o ponto de vista de investimento. Então, nós recentemente tivemos uma empresa de um ramo que não tem nada a ver com o ramo farmacêutico, nem químico, e que em função de uma matéria que saiu da mídia de um projeto que nós tínhamos de enfrentamento à covid, essa empresa entrou em contato conosco porque estava interessada em doar recursos para projetos que viessem a beneficiar à sociedade e nós conseguimos, olha, um excelente aporte financeiro, surpreendente, vindo de uma fonte que a gente não tinha a mínima ideia, como consequência dessa divulgação de dados. Então, nesse sentido, eu sou sim, estimulado a fazer isso que quanto mais gente, mais alunos procuram a gente, mais a gente se consolida como grupo de pesquisa regional, institucional, né?

Para PQu7, dados abertos de pesquisa podem contribuir para a divulgação científica, de modo a proporcionar maior interação entre o conhecimento gerado em universidades e sua aplicação na sociedade. Isso porque, essa receberia um retorno de todos os recursos que aplicou para o desenvolvimento de pesquisas.

De acordo com PQu4, a divulgação de dados abertos seria muito interessante para ampliar o conhecimento de pesquisas produzidas no país. Para tanto, PQu4 sugeriu que agências de fomento também tornassem públicos, por meio de plataformas específicas, os relatórios intermediários e as avaliações decorrentes de projetos financiados com dinheiro público. Sobre isso, PQu4 comentou:

Eu acharia também muito bom, assim, a agência de fomento, ela tem lá um lugar, eu vejo o nome dos trabalhos, mas eu não vejo os trabalhos depois. Ou seja, eu acho importante que também a gente tenha acesso diretamente aos relatórios, às coisas todas que os avaliadores postarem, mas isso acaba sendo restrito, somente à agência de fomento, hoje.

- Subcategoria “Recursos destinados à abertura de dados”

A partir dos relatos obtidos na pesquisa, principalmente os de PQu1, PQu4, PQu8 e PQu10, percebeu-se que ter em mãos recursos destinados à abertura de dados é um dos fatores que influenciam tal abertura no contexto de pesquisadores. Nessa direção, PQu10 ressaltou que o movimento da Ciência aberta é recente e pauta-se, principalmente, na alegação de que pesquisas financiadas com dinheiro público deveriam ser publicamente disponibilizadas. Entretanto, em algumas situações, a publicação de dados abertos de pesquisa exige o pagamento de custos relacionados, por exemplo, à manutenção de plataformas digitais de comunicação científica. Muitas vezes, esses custos são de responsabilidade de universidades que criam espécies de fundos para dar suporte a esse tipo de publicação, todavia, essa não é uma realidade amplamente vivenciada no contexto brasileiro. Sobre o aspecto mencionado, PQu10 fez o seguinte comentário:

Então, falando de ciência aberta, começa um movimento, talvez eu possa dizer nos últimos dez anos, ou talvez um pouco menos, de que esse conhecimento passa a ser público, como deveria ser, afinal de contas, no meu caso, quem paga pela maioria dos meus investimentos sou eu, são recursos públicos, ou recursos de empresas privadas via incentivo fiscal, de certa maneira, também é público. Agora, como fazer? Então, eu sou favorável, agora como fazer? Você precisa, por exemplo, pagar para que se mantenha no ar, no site, esse material público. Então, as universidades que eu colaboro na Europa, fora dela, elas criam fundos para poder suportar esse tipo de publicação, que há um custo. E o dólar, uns quinhentos dólares, dois mil, depende do tipo da revista, revistas convencionais cobram quatro mil dólares, e às vezes, até mais. Então, sintetizando a resposta, sou favorável, mas a gente precisa criar maneiras de subsidiar essa publicação aberta, e que no Brasil, ainda não existem.

De acordo com PQu1, é preciso que universidades disponham de infraestrutura tecnológica, física e técnica para lidar com a disponibilização de dados de pesquisa. Sendo assim, PQu1 levantou a problemática na qual a manutenção de repositórios técnicos é dispendiosa e a de que não haveria garantias efetivas de que dados seriam preservados a longo prazo.

Ressaltando o aspecto da infraestrutura tecnológica no contexto de abertura de dados, PQu8 sugeriu a criação de um repositório nacional de dados abertos que reunisse de modo sistemático, dados de diferentes universidades do país, como se pode observar em sua fala:

Eu acho que poderia haver, talvez, uma plataforma nacional que possa fazer isso, que nem a gente aqui, que é o repositório da universidade, mas só da universidade. A minha universidade vizinha não sei se faz isso, então, talvez, se houvesse um sistema nacional para você colocar os seus dados de

forma aberta, eu acho que isso ajudaria bastante. Assim, isso tem que ser de uma forma bem amigável também.

Considerando a relevância da disponibilização de recursos destinados à abertura de dados, PQu4 disse que a previsão do custeio da abertura de dados disposta em editais de financiamento à pesquisa é um fator importante para estimular cientistas a realizá-la.

4.3.5.2 Categoria “Fatores inibidores à abertura de dados”

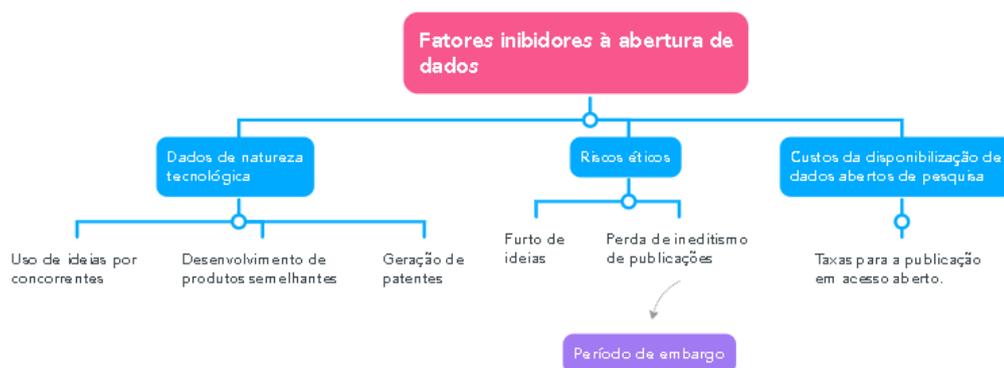
A categoria “Fatores inibidores à abertura de dados de pesquisa” compila os aspectos identificados no estudo que contribuem para desencorajar tal abertura. Assim sendo, trata-se do conjunto de forças que atuam sobre pesquisadores, de maneira a inibi-los em relação à abertura dos dados que geram.

Conforme é mostrado na Figura 47, esta categoria desenvolveu-se a partir de três subcategorias, sendo elas: 1) dados de natureza tecnológica; 2) riscos éticos; 3) custos da disponibilização de dados abertos de pesquisa. A primeira subcategoria compreende a preocupação com dados de natureza tecnológica sobre o uso de ideias por concorrentes e o desenvolvimento de produtos semelhantes por outras instituições. Além disso, ressalta-se que alguns dados não poderiam ser disponibilizados durante o andamento de pesquisas por potencializarem a geração de patentes.

A segunda subcategoria abrange riscos éticos, como o de furto de ideias e perda de ineditismo de publicações, devido à disponibilização antecipada de dados de pesquisa. Com a finalidade de mitigar tais riscos foi proposto que houvesse um período de embargo que antecederesse a disponibilização de dados abertos de pesquisa.

Finalmente, a última subcategoria refere-se aos custos da disponibilização de dados abertos de pesquisa, que podem incluir taxas onerosas para a publicação em acesso aberto. Entretanto, pesquisadores relataram que não recebem financiamento para essa finalidade.

Figura 47 – Categoria Fatores inibidores à abertura de dados de pesquisa - Química



Fonte: elaborado pela autora

- Subcategoria “Dados de natureza tecnológica”

De acordo com as informações obtidas no estudo, principalmente fornecidas por PQu2, PQu3, PQu5 e PQu10, dados de natureza tecnológica podem inibir a abertura de dados gerados por pesquisadores. Isso porque trata-se de uma categoria de dados que necessita de atenção especial para evitar o uso de ideias por concorrentes, e o consequente desenvolvimento de produtos semelhantes por outras instituições. Além disso, muitos desses dados podem ser fonte inicial para a geração de patentes, não sendo disponibilizá-los abertamente à ampla audiência.

Segundo PQu5, é fundamental que os riscos relacionados à exposição de dados de natureza tecnológica sejam mitigados. Isso, em razão da velocidade de produção científica, que é muito mais rápida no âmbito corporativo do que no ambiente acadêmico. Logo, quando dados dessa natureza são disponibilizados antecipadamente, os mesmos podem ser desenvolvidos com maior agilidade por outras instituições que, por sua vez, podem dispor de infraestrutura aprimorada para o desenvolvimento de pesquisas.

Semelhantemente, PQu10 afirmou que determinados dados com caráter de inovação tecnológica necessitam ser resguardados devido a pactos contratuais, previsão de estabelecimento de parcerias com instituições privadas sobre o desenvolvimento de produtos, ou por constituírem parte do processo de registro de patentes.

Reiterando o aspecto de que não seria cabível a divulgação de dados com natureza tecnológica, PQu3 explicou:

Então, eu acho que dados, alguns dados, eles podem ser abertos sem problema nenhum. Agora, outros dados não, como por exemplo, alguns dados que estão relacionados com a inovação tecnológica, alguns dados que estão relacionados com a forma que a pessoa trabalha, entendeu?

De mesmo modo, PQu2 comentou que o pesquisador que produziu dados é livre para explorar as possibilidades trazidas por esses em seus estudos, inclusive, negociando resultados obtidos com empresas de iniciativa privada. Em se tratando de projetos de transferência de tecnologia, PQu2 comentou que há um certo temor de que dados sejam acessados indevidamente, sendo copiados por indivíduos não autorizados.

No momento em que a gente conclui uma etapa importante de um projeto em que não há perspectiva desse projeto se tornar uma transferência de tecnologia, não há perspectiva desse projeto ser negociado com a iniciativa privada, por exemplo, nem há algum temor de que esse projeto seja copiado e terminado por outro e não por nós, a gente faz essa divulgação das formas mais abertas das quais eu já te falei antes, quer dizer, a gente divulga em congressos sejam eles locais, regionais, nacionais, internacional, a gente publica nas revistas.

- Subcategoria “Riscos éticos”

Por meio dos relatos de entrevistados, principalmente, PQu2, PQu3, PQu7, PQu8, PQu9 e PQu10, notou-se que a exposição de dados a riscos éticos é um fator que inibe a abertura de dados. Para mitigar tais riscos, PQu9 afirmou que solicita àqueles que tem interesse em utilizar dados de sua autoria a assinar um termo com condições específicas de uso de dados e de citação obrigatória do laboratório responsável pela geração dos dados.

Outro aspecto mencionado relativo à ética foi o risco de furto de ideias, o que, na visão de entrevistados, é altamente prejudicial para o andamento de pesquisas. Nesse sentido, PQu3 lembrou a possibilidade de pesquisadores usurparem a autoria de estudos realizados por outros cientistas:

Por que tudo isso pode ser [furto de ideias], a gente vê aí na mídia, né? Vários casos de pessoas, por exemplo, ultimamente tive um em relação à Covid lá, alguns trabalhos em que o cara pegou dados do outro e publicou, acochambrou para cá e fez aquilo. Isso aí acontece, entendeu? Isso aí vai acontecer, e acontece não só na questão da covid, mas em todo nível de pesquisa isso pode acontecer, das pessoas pegarem os dados, entendeu?

De modo semelhante, PQu7 afirmou que o risco de furto de ideias é o maior problema relacionado à abertura de dados de pesquisa, como foi exposto em sua fala:

Eu acho que o maior problema é o furto de ideais, e entre as próprias [pessoas] da área. E não é nem, às vezes é intencional e às vezes não. Você lê o e-mail e você diz: 'não, eu posso fazer isso', só que daí a pessoa que começou aquele trabalho, às vezes não tem a mesma velocidade daquele que pegou pelo caminho, pela metade do caminho. Então, às vezes, é isso que faz com que o pesquisador não deixe tão aberto o trabalho que está desenvolvendo.

Com relação ao aspecto ético, PQu8 ressaltou que dados poderiam ser disponibilizados, contanto que participantes de pesquisas não sejam identificados e que esses dados não necessitem de sigilo por motivo relativo à pesquisa. Já PQu10, destacou que o trabalho com determinado conjunto de dados deve ser resguardado antes de ser disponibilizado, uma vez que há a possibilidade de geração de patentes.

Outro ponto relacionado à questão ética é a perda de ineditismo de publicações, ocasionada pela disponibilização antecipada de dados de pesquisa. Segundo PQu9, isso pode afetar o aceite de dados e artigos em periódicos científicos, na medida em que tal conteúdo perderia um certo valor científico, por se tratar de dados que já teriam sido publicados abertamente. Sob tal perspectiva, alguns periódicos ficariam desinteressados nesse tipo de publicação.

Segundo PQu2, a perda do ineditismo de descobertas científicas é um risco quando se disponibiliza dados de pesquisa de estudos ainda em andamento. Nessa direção, PQu2 destacou o desequilíbrio no processo de produção científica, no qual o cientista que tem a seu dispor maior infraestrutura de pesquisa poderia reutilizar dados e publicar os resultados obtidos, antes mesmo do pesquisador que os produziu. Sobre essa desvantagem, PQu2 comentou:

Se eu for divulgar um resultado de uma pesquisa que não está fechada ainda, mas ela é interessante, eu abro a possibilidade. Se ela não estiver protegida de um cientista no primeiro mundo terminar aquilo que eu levaria dois, três anos, para terminar em seis meses, ele pode pegar minha ideia, a partir do ponto que eu publiquei. Ele não está copiando nada, mas ele vai estar pegando a ideia, vai estar pegando os resultados que foram importantes e aquilo que eu teria que fazer à frente, mas que eu levaria muito mais tempo, porque eu estou em desvantagem em termos de infraestrutura, eu estou em desvantagem em termos de recursos e lá eles conseguem fazer isso em um prazo muito menor. Não é à toa que a gente tem tantos produtos nossos, produtos naturais, matéria primas, que saem do país e rapidamente retornam como produto acabado, porque lá fora eles têm uma capacidade de trabalhar muito mais rápida do que nós.

Para PQu3, a perda do ineditismo devido à publicação antecipada de dados é algo muito sério no contexto de inovação tecnológica. Isso porque outros cientistas

poderiam concluir os estudos originais e publicar esses dados antes mesmo dos responsáveis por sua criação. Nessa direção, PQu3 comentou:

Se o meu colega da universidade está trabalhando com uma determinada planta, eu não vou ficar trabalhando naquela mesma planta, naquele mesmo objetivo lá. Eu não, eu não vou, eu vou pegar outras plantas, eu respeito.

Com o propósito de mitigar riscos relativos à perda do ineditismo ocasionada pela publicação antecipada de dados de pesquisa, PQu9 sugeriu o cumprimento de um período de embargo que antecederesse a disponibilização de dados abertos de pesquisa, assim como ocorre no registro de patentes de processos e produtos:

É basicamente isso que eu coloquei, eu acho que devia existir uma proteção por um tempo para os dados. Os dados brutos, eu acho que eles teriam que ser posse da universidade, por exemplo, eu produzo um espectro de ressonância magnética nuclear, e ele vai para esse banco de dados, a interpretação que eu faço, a publicação que eu faço com base naquilo ali é responsabilidade minha, mas se a revista ou o livro quiser olhar o dado bruto, a universidade pode conceder essa vitória para a revista. Porém, o dado trabalhado, o resultado da pesquisa, a interpretação intelectual que vai originar uma hipótese, que vai comprovar uma hipótese, uma teoria, ele deveria ser protegido por um tempo e depois ser aberto, como o modelo que se usa para o patenteamento de processos e produtos.

- Subcategoria “Custos da disponibilização de dados abertos de pesquisa”

De acordo com informações obtidas neste estudo, principalmente a partir dos relatos de PQu4, PQu6 e PQu9, os custos envolvidos na disponibilização de dados abertos de pesquisa constituem um fator de inibição à abertura dos mesmos.

Apesar de ter interesse em compartilhar dados abertos de pesquisa, PQu4 afirmou que não dispõe de condições financeiras para fazê-lo, devido aos altos custos para a publicação em acesso aberto. Isso porque não recebe nenhuma verba extra para financiar esse tipo de publicação, o que também não é previsto nos editais de financiamento os quais costuma pleitear por recursos. Sobre isso, PQu4 comentou:

E aí, isso tudo é custo, gostaria de publicar abertos, né? Só que eu não tenho condições. O meu salário não dá para pagar a publicação dos meus dados, então, eu fico restrito ao próprio pagamento, né? E disponibilizar isso, porque é custo, eu não tenho custeio para poder fazer isso [...] os nossos editais não contemplam esse tipo de abertura, também de dados. Então, é orçamento, a verdade é essa. Eu preciso de um orçamento para poder disponibilizar esses dados de forma aberta.

Além disso, PQu4 destacou que o custo de produção de dados de pesquisa já é muito alto, isso, sem mencionar os custos destinados à publicação de conteúdos em acesso aberto. Tal situação é percebida por PQu4, em especial, no que se refere à aquisição de

reagentes e equipamentos para a realização de experimentos. Sendo assim, percebe-se a importância de recursos direcionados, especificamente, à publicação de dados abertos de pesquisa.

De maneira semelhante, PQu6 destacou que as taxas para a publicação de conteúdos em acesso aberto são onerosas, não cabendo ao pesquisador arcá-las em sua totalidade. Por essa razão, PQu6 sugeriu que o custeio relacionado a esse tipo de publicação fosse de responsabilidade da instituição ao qual pertence o pesquisador ou da agência de fomento que financiou o estudo realizado. Nesse sentido, o custeio institucional poderia estimular a abertura de dados de pesquisa dentre os cientistas da área.

Com base no exposto, é cabível a reflexão sobre o que propôs PQu9 a respeito da criação de um programa encabeçado por uma entidade de representação mundial que promovesse um consórcio de dados entre membros da comunidade científica. Tal consórcio poderia abranger conjuntos de dados abertos de diferentes países, nos quais os custos envolvidos na publicação desses dados seriam providos pelo programa estabelecido.

De outro vértice, e semelhantemente ao que foi verificado no contexto da Sociologia, o estudo identificou que uma eventual obrigatoriedade de depósito de dados abertos de pesquisa estabelecida por instituições científicas e acadêmicas foi vista como um fator que poderia estimular tal abertura, bem como inibi-la. Por isso, este trabalho considerou a criação de duas subcategorias que agrupam, em cada uma delas, argumentos favoráveis ou desfavoráveis à obrigatoriedade mencionada.

Dessa maneira, a opinião favorável à obrigatoriedade de depósito de dados como fator que estimula a abertura de dados de pesquisa está representada na categoria: “Favorável à obrigatoriedade”. De outro modo, a opinião desfavorável a essa obrigatoriedade, considerando-a um aspecto inibidor à abertura de dados de pesquisa, está representada na categoria: “Desfavorável à obrigatoriedade”.

4.3.5.3 Categoria “Favorável à obrigatoriedade”

Conforme é apresentado na Figura 48, a categoria “Favorável à obrigatoriedade” foi desenvolvida a partir da subcategoria “Pesquisas fomentadas com dinheiro público” e “Comprovação de dados por editores científicos”.

A primeira sustenta-se no argumento de que pesquisas financiadas por meio de recursos públicos deveriam ter seus dados abertamente disponíveis, uma vez que a sociedade estaria recebendo um retorno daquilo que investiu. Já a segunda subcategoria, destaca que a obrigatoriedade de disponibilização de dados abertos de pesquisa é fundamental para a comprovação de dados por editores científicos e consequente ateste da veracidade de informações veiculadas em determinado estudo.

Figura 48 – Categoria Favorável à obrigatoriedade - Química



Fonte: elaborado pela autora

- Subcategoria “Pesquisas fomentadas com dinheiro público”

Com base nas informações obtidas no estudo, principalmente a partir dos relatos de PQu7 e PQu8, verificou-se que aqueles favoráveis à obrigatoriedade da disponibilização de dados abertos de pesquisa sustentaram-se no argumento central de que pesquisas fomentadas com dinheiro público deveriam ter os dados obtidos disponíveis abertamente.

De acordo com PQu8, a sociedade deveria receber o retorno daquilo que investiu no desenvolvimento científico, por isso, dados produzidos mediante financiamento público também deveriam ser públicos. Nessa direção, PQu8 chegou a

propor a exigência da abertura de dados de pesquisa para o recebimento de financiamento concedido por agências de fomento.

Ao discorrer sobre a possibilidade de a universidade em que atua e de agências de fomento tornarem obrigatório o depósito de dados abertos de pesquisa, PQu6 reiterou o aspecto de dados que foram gerados mediante auxílio financeiro público, que possibilitou a realização do estudo. Para PQu6, essa noção deveria partir do pesquisador, que teria como propósito, a divulgação do conhecimento científico, além do sentimento de prestação de contas à sociedade.

Eu acho ótimo que seja obrigatório. Na verdade, eu acho que teria que partir do próprio pesquisador. Você não faz ciência só para você, para o seu laboratório e para um nicho específico, você tem que mostrar para a sociedade como o dinheiro está sendo gasto [...] assim, eu considero a agência de fomento pagando a pesquisa e a universidade pública pagando o pesquisador. Na verdade, eles estão financiando como um todo, pesquisador e pesquisa. Então, eu não vejo diferença, eu acho que tem que ter obrigatoriedade.

Semelhantemente, PQu7 trouxe à tona o aspecto da prestação de contas de cientistas à sociedade sobre investimentos em pesquisas realizadas. De acordo com PQu7, “o que é financiado tem que ser aberto”. Portanto, não somente o documento final de pesquisas deveria ser disponibilizado após sua conclusão, mas também dados e relatórios intermediários. Ademais, esses deveriam ser amplamente conhecidos, e não ficar concentrados nas mãos de avaliadores de projetos financiados por agências de fomento. Sob tal perspectiva, muitos dados estariam inacessíveis ou apresentariam inúmeras restrições até que fossem disponibilizados abertamente.

- Subcategoria “Comprovação de dados por editores científicos”

Com base nas informações obtidas no estudo, em especial, dispostas nos relatos de PQu2, PQu5 e PQu6, a possibilidade da exigência do depósito de dados a pesquisadores foi percebida como um estímulo para a comprovação de dados, principalmente, no âmbito de editores científicos. Nesse sentido, PQu2 destacou que dados provenientes de estudos não concluídos não poderiam ser disponibilizados, haja vista que os impactos dos mesmos ainda não teriam sido mensurados e, possivelmente, poderiam sofrer alterações conforme o andamento da pesquisa.

Para PQu5, a comprovação da veracidade de dados é fundamental para dar transparência aos trabalhos de pesquisadores, sendo uma maneira de proteger eticamente universidades e agências de fomento, bem como dar à sociedade um retorno sobre o conhecimento que está sendo produzido. Sobre isso, PQu5 comentou:

É uma necessidade porque você protege aqueles que trabalham de maneira correta, que não tem dado a esconder. Você protege a instituição, protege as agências que estão financiando aquele tipo de pesquisa e, sobretudo, você dá um retorno à população. Então, a transparência em qualquer nível, situação de processo, ela é sempre bem-vinda.

De acordo com PQu6, uma possível obrigatoriedade de depósito de dados abertos de pesquisa exigida por editores científicos é fundamental para verificar a idoneidade da pesquisa, e, conseqüentemente, comprovar a fidedignidade da origem de dados. PQu6 lembrou a existência de periódicos predatórios, os quais podem aceitar qualquer trabalho submetido, sem a realização do devido trâmite de revisão por pares, somente com a finalidade de aumentar o número de trabalhos publicados.

[Entrevistadora]: E no caso dos editores científicos, qual é a sua opinião sobre tornar obrigatório o depósito dos dados?

Concordo, porque afinal, o editor que aceita o artigo para publicar, ele quer comprovar que seja idôneo, então, não adianta você só escrever porque tem muito isso em revistas predatórias. Então, eles estão tão desesperados, algumas revistas estão tão desesperadas para receber, porque é open access que aceita qualquer tipo de trabalho e trabalho, muitas vezes, que não foi revisado por pares e com rigor. Então, eu acho que o editor tem esse papel. 'Tá, eu aceito o seu artigo, porque no artigo você faz uma leitura linda e maravilhosa, porque você vai destacar o que foi e o que deu certo. Tá, cadê os dados brutos? Cadê os espectros? Cadê tudo que comprova isso daqui?' Isso tem que ser avaliado pelos pares também. Então, é assim, muitas revistas já pedem como material suplementar, outras não pedem, a gente já instituiu, como sempre a gente prepara esse material suplementar porque não cabe para ser publicado, é um material que fica como um link para você ter acesso. Mas ele está na revista, só não está na revista [juntamente] com os artigos, um atrás do outro. Está lá com o link que você vai acessar depois. Mas eu acho que todos os artigos tinham que ter, porque na nossa área, essencialmente, a gente precisa mostrar esses dados até para ser avaliado pelos pares, quando você submete o artigo.

4.3.5.4 Categoria “Desfavorável à obrigatoriedade”

Assim como é mostrado na Figura 49, a categoria “Desfavorável à obrigatoriedade” pode ser entendida por meio da subcategoria “Concessão voluntária”. Essa subcategoria baseia-se no argumento de que a abertura de dados deve ser fruto de uma escolha consciente do pesquisador. Sendo assim, caberia ao mesmo disponibilizá-los mediante o estabelecimento de acordos com agências de fomento e editores científicos. Além disso, o termo “obrigatório” pode apresentar uma conotação

negativa, na qual pode estar associado a algo imposto. Consequentemente, pesquisadores poderiam não aceitar a ideia facilmente.

Figura 49 – Categoria Desfavorável à obrigatoriedade - Química



Fonte: elaborado pela autora

- Subcategoria “Concessão voluntária”

Com base nos relatos obtidos no estudo, principalmente, os de PQu2, PQu3 e PQu5, o depósito de dados abertos não deveria ser obrigatório por instituições científicas e acadêmicas, mas sim, resultado de uma escolha consciente do pesquisador em relação a essa abertura. Ou seja, trata-se de uma concessão voluntária de dados.

De acordo com PQu3, apesar de pesquisas serem financiadas por meio de recursos públicos, elas integram a produção intelectual do pesquisador e também dizem respeito às suas particularidades como autor. Sob tal perspectiva, a abertura de dados seria uma atitude de competência do pesquisador que os gerou, conforme pode-se observar no diálogo entre entrevistadora e PQu3:

[Entrevistadora]: Então, no primeiro momento, é a sua opinião sobre a possibilidade de a universidade em que você atua tornar esse depósito obrigatório de forma aberta, como você enxerga isso? Eu sou contra, terminantemente contra, né? [...] é como teu CPF, né? Como tua conta bancária. Então, porque a gente não deixa nossa conta toda aberta? É uma coisa muito nossa, e a pesquisa não, né?

A pesquisa é sim, apesar de eu trabalhar em uma instituição pública. Eu devo satisfação à sociedade. Mas eu devo satisfação antes da sociedade, eu devo a mim mesmo, entendeu?

[Entrevistadora]: E a questão dos editores científicos? Nesse caso de obrigatoriedade de abertura de dados?

A abertura de dados não, a abertura de dados, não! A gente pode depositar os dados, mas a abertura, ela tem que ser com uma concessão nossa final. A gente conceder, ou não, o que tem que ser. Por que eu acho que tudo tem que ser definido aí.

De modo semelhante, PQu5 destacou o aspecto de que o termo “obrigatório” geralmente possui uma conotação negativa, na qual a primeira reação obtida é a não aceitação por algo que foi imposto. Sendo assim, é bastante provável que pesquisadores reivindiquem qualquer imposição nesse sentido.

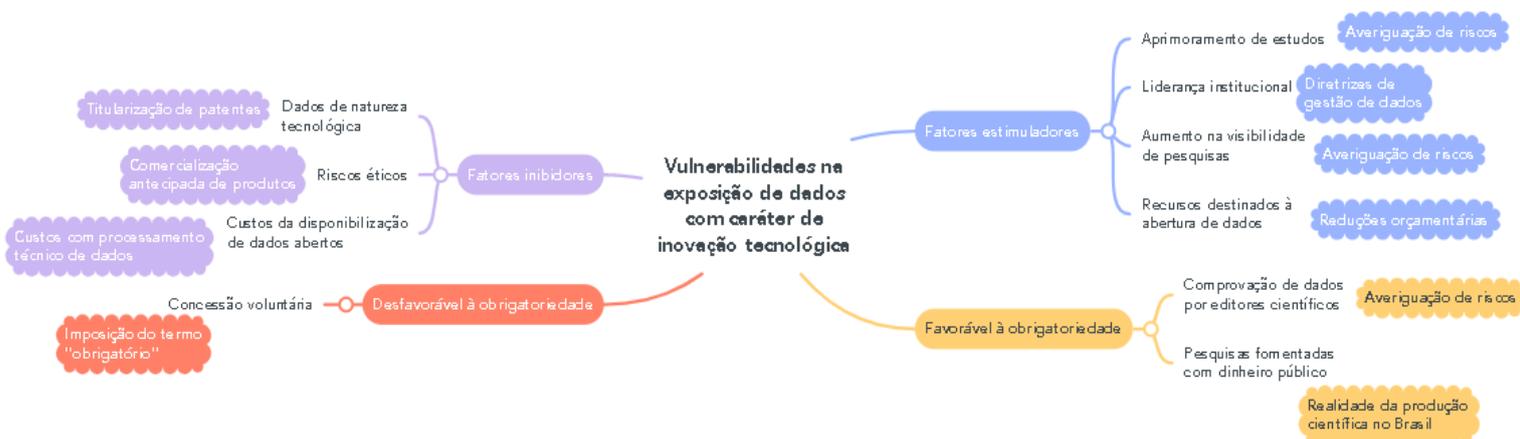
Além disso, PQu5 comentou que caso de a obrigatoriedade de depósito de dados abertos for exigida por agências de fomento, caberia ao pesquisador aceitar, ou não, os recursos obtidos para o financiamento da pesquisa, uma vez que o mesmo poderia conceder abertamente, ou não, os dados gerados na pesquisa.

Assim também, PQu2 ressaltou que nas relações entre pesquisador e editores científicos há acordos contratuais nos quais, geralmente, o primeiro concede direitos autorais relativos a conteúdos científicos aos segundos. Logo, presume-se uma concessão voluntária do pesquisador, tal como deveria ocorrer em relação aos dados de suas pesquisas. A respeito disso, PQu2 comentou: “no momento em que eu coloquei o material na mão da editora, ela pode divulgar aquilo amplamente, desde que preservada a autoria, obviamente, daquela informação”.

4.3.6 Codificação seletiva e teoria substantiva referente aos fatores que influenciam a abertura de dados de pesquisa em Química

No que se refere à área da Química, dados obtidos no estudo apuraram a centralidade da categoria “Vulnerabilidades na exposição de dados com caráter de inovação tecnológica” nas relações estabelecidas com as demais categorias identificadas. Assim como é representado na Figura 50, pode-se destacar tais relações da seguinte maneira:

Figura 50 – Vulnerabilidades na exposição de dados com caráter de inovação tecnológica e sua relação com fatores que influenciam a abertura de dados: Química



Fonte: elaborado pela autora

- para que dados sejam disponibilizados abertamente, constituindo-se insumos para o aprimoramento de pesquisas na área de Química, comprovação de informações a editores científicos e para o aumento na visibilidade de estudos, é indispensável averiguar se há vulnerabilidades na exposição de dados com caráter tecnológico, os quais não poderiam ser expostos de qualquer maneira;
- muitos dados com caráter de inovação tecnológica constituem-se ponto de partida para se pleitear a titularização de patentes, não podendo ser amplamente disponíveis por determinado período;
- dados com caráter de inovação tecnológica podem estar atrelados a pactos contratuais para o desenvolvimento de novos produtos. Logo, necessitam de proteção para não correrem o risco ético de uma comercialização antecipada;

- para que dados abertos sejam amplamente acessíveis, há custos envolvidos, tais como o processamento técnico de conjuntos de dados com caráter tecnológico, visando minimizar vulnerabilidades em sua exposição;
- pesquisas brasileiras caracterizam-se, sobretudo, pelo fomento oriundo de instituições públicas, que, por sua vez, tem sido alvo de constantes reduções orçamentárias que impactam diretamente a disponibilização de dados abertos;
- a existência de uma liderança institucional contribuiria para minimizar riscos relacionados à vulnerabilidade na abertura de dados com caráter de inovação tecnológica, por meio do fortalecimento de discussões sobre o tema e do estabelecimento de diretrizes de gestão de dados abertos de pesquisa;
- considerando vulnerabilidades na abertura de dados com caráter de inovação tecnológica, em geral, pesquisadores da área de Química se posicionam contrariamente a uma possível obrigatoriedade de disponibilização de dados exigida por instituições científicas e acadêmicas. Isso porque o termo “obrigatório” pode ser vinculado negativamente a uma imposição, na qual pesquisadores não se sentem nada confortáveis.

À luz da literatura científica, a seção seguinte abordará a categoria central – vulnerabilidades na exposição de dados com caráter de inovação tecnológica – segundo as relações estabelecidas com as demais categorias identificadas no estudo, que, por seu turno, dizem respeito a fatores que influenciam a abertura de dados de pesquisa.

4.3.6.1 Vulnerabilidades na exposição de dados com caráter de inovação tecnológica e sua relação com os fatores que influenciam a abertura de dados

O estudo constatou que a vulnerabilidade na exposição de dados com caráter de inovação tecnológica é o fator predominante que influencia a abertura de dados dentre os entrevistados da área de Química. Geralmente, dados dessa natureza requerem tratamento diferenciado com a finalidade de minimizar o risco do uso de ideias por concorrentes e, por conseguinte, a criação de produtos similares por outras instituições.

Também se destaca que muitos dados constituem insumos para se pleitear a titularização patentes, não sendo possível ser disponibilizá-los abertamente. Associado a essa questão, está o aspecto abordado por pesquisadores de que a velocidade de produção científica no contexto de empresas privadas é bem mais veloz do que no ambiente acadêmico. Isso porque, em geral, tais instituições dispõem de infraestrutura avançada para o desenvolvimento de suas pesquisas.

Além disso, dados com caráter de inovação tecnológica podem estar arrolados a pactos contratuais resultantes do estabelecimento de parcerias público-privadas para o desenvolvimento de pesquisas, e, especificamente, de algum produto. Portanto, é fundamental que esse tipo de dado não seja exposto a riscos éticos, devendo ser protegido e amparado por vias legais que explicitem os termos pelos quais o mesmo possa ser disponibilizado.

De acordo com Ferreira, Guimarães e Contador (2009), o desenvolvimento de ações inovativas no campo tecnológico é importante para induzir um país ao crescimento econômico. Mas para isso, é preciso que haja garantias jurídicas contra imitadores que exploram ilegalmente inovações tecnológicas. Um dos instrumentos utilizados por pesquisadores para evitar tal exploração e preservar o seu direito de propriedade intelectual é o registro de patentes.

Segundo a Coordenação de Inovação Tecnológica (CIT) da Universidade Federal de Pelotas – UFPel (2018), uma patente é uma titularização temporária, concedida a inventores pelo Estado, que atesta os direitos autorais sobre determinado produto ou processo, tal como é apresentado na seguinte definição:

***PATENTE:** É um título de propriedade temporário, concedido pelo Estado, aos inventores ou empresas que passam a possuir os direitos sobre a invenção, seja ela relativa a um produto, a um processo de fabricação ou ao aperfeiçoamento de produtos e processos preexistentes, como recompensa aos esforços despendidos nessa criação. Com a posse da Carta-Patente, o titular tem a exclusividade de exploração de seu produto, podendo industrializar, vender ou transferir a terceiros, definitivamente ou temporariamente os seus direitos. Desta forma, a patente permite a seu detentor uma reserva de mercado por tempo determinado. Terminado o prazo do privilégio concedido, a criação industrializável protegida cai em domínio público (CIT, 2018, p. 7).*

Nesse contexto, um grande dilema se apresenta ao pesquisador, que geralmente opta por duas opções: publicar dados e resultados obtidos em um artigo científico, ou registrar um pedido de patente de Invenção. Essa é uma decisão que impacta a cadeia produtiva de pesquisa no ambiente acadêmico, e inclusive, no direcionamento

de políticas públicas nesta área, uma vez que programas de pós-graduação são periodicamente avaliados quanto à produção científica que apresentam (CIT, 2018).

Logo, programas mais produtivos tendem a atrair maiores financiamentos à pesquisa. Nesse contexto, é imprescindível que pesquisadores tenham total clareza sobre as implicações ocasionadas pelo registro de patentes e pela publicação de artigos, para decidir qual caminho tomarão com os dados de sua responsabilidade autoral (CIT, 2018). Como se pode observar, a Figura 51 diferencia em vários aspectos o registro de patentes e a publicação de artigos:

Figura 51 – Diferenças entre o registro de patentes e a publicação de artigos



Fonte: Adaptado de CIT, 2018, p. 6

Excluídos ou mitigados os riscos relacionados à vulnerabilidade na exposição de dados com caráter de inovação tecnológica, e quando for eticamente possível disponibilizá-los (sem prejuízo ao sigilo e com tratamento técnico e proteção adequados), os mesmos podem trazer inúmeros benefícios à produção científica. De

acordo com Jomier (2017), dados abertos podem favorecer a reprodutibilidade de pesquisas e, conseqüentemente, reduzir o tempo empregado na realização de experimentos já realizados.

Somando-se aos benefícios apontados por Jomier (2017), podem ser destacados outros três: 1) o aprimoramento de pesquisas na área, a partir de dados reutilizados; 2) o aumento na visibilidade de estudos, os quais podem desencadear novos ciclos de produção científica; 3) comprovação de dados por editores científicos, de modo a atestar a veracidade de informações veiculadas em artigos científicos.

As vantagens proporcionadas pelo acesso aberto a publicações científicas não se esgotam aí. De acordo com Melero (2005), esse tipo de acesso está relacionado ao rompimento de restrições ao conhecimento científico, de modo a aumentar a visibilidade de pesquisas e disseminá-las à sociedade, principalmente quando se tratar de estudos financiados com recursos públicos. Sendo assim, o acesso aberto favorece o fluxo de dados no sistema tradicional de comunicação científica, sobretudo, devido a novas funcionalidades provenientes do ambiente da Internet.

Todavia, por mais que a disponibilização de dados abertos permita o seu acesso e compartilhamento de forma gratuita à ampla audiência, há um custo envolvido em todo o processo relacionado aos dados, principalmente financeiro e de tempo. Esse custo pode incluir o processamento técnico de conjuntos de dados, a manutenção de repositórios institucionais e até o pagamento para editores científicos relativos à publicação em acesso aberto.

De acordo com Jomier (2017) a disponibilidade de dados de pesquisa exige o cumprimento de alguns requisitos para ser desempenhada efetivamente, como o manuseio de dados gerados em diferentes áreas do conhecimento, a infraestrutura tecnológica robusta para depositar grandes quantidades de dados e o conhecimento sobre licenças de compartilhamento e uso de dados de pesquisa.

Segundo Ferreira Guimarães e Contador (2009), sempre que uma instituição decide instaurar processos de inovação tecnológica de forma deliberada e organizada, por meio da implantação ou manutenção de um departamento ou centro tecnológico, essa decisão resultará em aumento de seus gastos. Por outro prisma, estima-se que os custos mencionados sejam convertidos de alguma maneira em benefícios futuros.

Com base no exposto, percebe-se a importância da destinação de recursos não somente à realização de pesquisas, mas também à abertura de dados gerados. Tal como apontou o presente estudo, já se sabe que pesquisas brasileiras são amplamente fomentadas com dinheiro público. Entretanto, nota-se também uma constante redução orçamentária no que se refere a investimentos na área de educação e tecnologia, o que impacta diretamente na disponibilização de dados abertos. Tal redução foi constatada no relatório do Observatório do Conhecimento sobre o balanço orçamentário brasileiro do ano de 2021, elaborado pelos economistas Silva e Zelesco (2022). Nele, pode-se destacar os seguintes pontos:

- em 2015, a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes) tinha um orçamento liquidado de R\$ 10,3 bilhões que passou para R\$ 3 bilhões, em 2021;
- o gasto discricionário destinado à educação superior foi de R\$ 14,49 bilhões em 2014, atingiu um pico de R\$ 15,67 bilhões em 2015, tendo sido reduzido desde então, até chegar em R\$ 5,5 bilhões em 2021;
- em 2014, os gastos com universidades e institutos federais foram aproximadamente R\$ 8,46 bilhões em 2014, passando para R\$ 3,5 bilhões em 2021, o que corresponde a quase 41,5% do valor inicial dessa série;
- somente 55,44% do que foi previsto em orçamento na Lei Orçamentária de 2021 foi efetivamente executado para as universidades.

De acordo com matéria publicada pelo Sindicato nacional dos docentes das instituições de ensino superior – Andes (2022)⁸³, as universidades e institutos públicos de pesquisa ocupam lugar central na produção do conhecimento científico no país. Entretanto, é cada vez mais notória a redução de investimentos destinados à área de Ciência e Tecnologia. Por conseguinte, pesquisadores vivenciam dificuldades que incluem a disponibilização insuficiente de recursos necessários à manutenção de laboratórios, equipamentos e à concessão de bolsas de estudo para o desenvolvimento de projetos de pesquisa.

⁸³ Disponível em: <https://www.andes.org.br/conteudos/noticia/pesquisa-nacional-luta-para-sobreviver-asfixiada-por-cortes-orcamentarios1>

Considerando a instabilidade resultante dessa avalanche de reduções no contexto científico brasileiro, jovens pesquisadores acabam sendo desestimulados a seguir a carreira acadêmica. Nesse sentido, destaca-se que a denominada “fuga de cérebros” pode aumentar exponencialmente nos próximos anos, caso o cenário de desestímulo permaneça (ANDES, 2022).

Outro aspecto importante para minimizar os riscos relacionados à vulnerabilidade na abertura de dados com caráter de inovação tecnológica é a presença de uma liderança institucional responsável pelos processos relativos à abertura e gestão de dados de pesquisa. Neste estudo, observou-se que pesquisadores ressaltaram a importância do fortalecimento de discussões sobre o tema e da definição de diretrizes institucionais que os orientem sobre como devem proceder mediante tal abertura.

À vista do exposto, destaca-se que bibliotecas estão cada vez mais se aproximando do contexto de dados de pesquisa e, em razão disso, começam a oferecer serviços de gestão de dados a seus usuários. Em um contexto universitário, Barbrow, Brush e Goldman (2017), afirmam que a gestão de dados se refere a um conjunto de ações que devem ser realizadas durante todo o ciclo da pesquisa científica. Nessa direção, cabe às bibliotecas dar suporte a docentes e pesquisadores na elaboração de planos de gestão de dados (com frequência demandados por agências de fomento à pesquisa), bem como viabilizar o compartilhamento e arquivamento adequado de dados.

Semelhantemente, Thomas e Urban (2018) explicam que serviços de dados desenvolvidos por bibliotecas acadêmicas têm se intensificado devido ao crescimento da quantidade de dados produzidos, à exigência feita por agências de fomento sobre a disponibilidade de dados de pesquisa e a consolidação de redes de compartilhamento de dados.

Ao falar a respeito de bibliotecas que desenvolvem serviços de gestão de dados, Silva (2016) destaca que é preciso considerar o ciclo de vida de dados no decorrer da pesquisa, o que abre margem para muitas possibilidades de interação entre bibliotecários e pesquisadores, desde a obtenção de dados. Nessa etapa, o bibliotecário pode prestar auxílio ao pesquisador na elaboração do Plano de gestão de dados, no uso de padrões de metadados e na seleção de repositórios para

submissão dos dados que produziu. Após serem coletados, esses dados receberão do bibliotecário tratamento técnico, de maneira que serão devidamente transcritos e armazenados em repositórios. No que se refere à etapa de preservação de dados, essa abrange procedimentos como a migração de dados e realização de *backups*. Bibliotecários são profissionais capazes de orientar pesquisadores em relação a formatos de preservação de dados, até mesmo migrando-os para plataformas mais atuais (SILVA, 2016).

Nesse contexto, destaca-se a função exercida pelos chamados “bibliotecários de dados”. Segundo Chiware (2020), esses são profissionais que se dedicam à gestão de dados de pesquisa no contexto de bibliotecas. Lançam mão de vários recursos informacionais e dão suporte a pesquisadores durante as atividades inerentes ao ciclo de vida dos dados.

Na perspectiva de Semeler e Pinto (2019), é imprescindível que o bibliotecário de dados demonstre suas habilidades tradicionais para prestar serviços condizentes com uma nova realidade, na qual é gerada uma grande quantidade de dados em curto espaço de tempo. Sendo assim, é fundamental que esse profissional se aproprie de conhecimentos relacionados a regulamentos de repositórios de dados de pesquisa, documentação, tratamento técnico, preservação e análise de dados, além de dispor ao seu alcance informações relativas ao uso de ferramentas de acesso a dados.

De outro vértice, Semeler, Pinto e Rozados (2019) explicam que é altamente recomendável que o bibliotecário de dados compreenda aspectos relacionados a linguagens de programação de computadores, uma vez que esse pode trabalhar em diferentes plataformas para a recuperação de informações e gestão de dados de pesquisa.

Por fim, considerando a vulnerabilidade na exposição de dados com caráter de inovação tecnológica um fator predominante de influência à abertura de dados, muitos entrevistados neste estudo se posicionaram contrariamente a uma possível obrigatoriedade de disponibilização de dados abertos demandada por instituições científicas e acadêmicas. Isso, embasado no argumento da escolha consciente do pesquisador em querer disponibilizar abertamente os dados que produz, ou não, o que também inclui aceitar ou recusar acordos estabelecidos com agências de fomento

e editores científicos. Ressalta-se, ainda, que o termo “obrigatório” pode ser interpretado negativamente por pesquisadores por estar associado a uma imposição não almejada.

5 DIRETRIZES PARA POLÍTICAS INSTITUCIONAIS DE GESTÃO DE DADOS ABERTOS DE PESQUISA EM UNIVERSIDADES

Com base nas falas dos pesquisadores entrevistados no estudo, cujo roteiro de entrevista foi elaborado a partir de definições e conceitos extraídos da literatura científica e técnica, foi possível obter percepções relacionadas a práticas de gestão de dados de pesquisa e a fatores que influenciam a abertura de dados. Só então, sob a perspectiva desses pesquisadores, pode-se elaborar um conjunto de diretrizes para políticas institucionais de gestão de dados abertos de pesquisa no contexto de universidades. Tais diretrizes são:

- solicitar ao pesquisador a elaboração de um Plano de gestão de dados em projetos de pesquisa desenvolvidos no âmbito da universidade;
- prezar pela proteção ética de dados de pesquisa produzidos na instituição;
- oferecer suporte a pesquisadores para a construção de repositórios/bancos de dados, como alternativa para o compartilhamento e depósito de dados gerados em projetos de pesquisa;
- utilizar metabuscadores para recuperar e integrar conjuntos de dados e metadados depositados por pesquisadores em outros repositórios/bancos de dados;
- adotar estratégias de *marketing* para promover o reuso de dados abertos disponíveis em repositórios e dar visibilidade a pesquisas cujos dados encontram-se fechados, por meio da divulgação de seus metadados;
- adotar estratégias de preservação de dados a longo prazo;
- viabilizar o reuso de dados abertos de pesquisa, prezando pelos direitos autorais sobre dados;
- elaborar uma Política de gestão de dados abertos de pesquisa;
- implantar serviço de gestão de dados de pesquisa;
- capacitar equipes de bibliotecários para a atuação em serviços de gestão de dados de pesquisa;
- fomentar discussões sobre a gestão de dados abertos de pesquisas na comunidade acadêmica;

- implantar sistema de recompensas para a valorização do compartilhamento de dados abertos de pesquisa;
- gerir recursos não somente para a divulgação de resultados de pesquisa publicados, mas também de dados de pesquisa produzidos na instituição;
- gerir dados abertos de pesquisa de modo a considerar diferenças disciplinares;
- promover o depósito de dados abertos de pesquisa em repositórios, por meio de vias dialógicas.

Convém ressaltar que se trata de orientações que, ao mesmo tempo, pretendem ser genéricas, visando sua aplicação em contextos disciplinares diversificados, e específicas, de modo a considerar as particularidades das áreas do conhecimento em relação aos dados de pesquisa. Além disso, contemplam a perspectiva de dados abertos de pesquisa e dos fatores relacionados a sua abertura, bem como pode ser observado no objetivo geral declarado no estudo.

Destaca-se também que essas diretrizes não se esgotam no presente trabalho. Ao contrário, abrem margem para novas reflexões sobre o tema, não somente no âmbito de Sociologia, ou Química, mas em um contexto mais amplo, ocupado pelas grandes áreas das Ciências humanas e Sociais, além das Ciências naturais. Dado o exposto, cada uma das diretrizes elaboradas será descrita nos tópicos seguintes, de maneira que se recomenda às universidades:

5.1 Solicitar ao pesquisador a elaboração de um Plano de gestão de dados em projetos de pesquisa desenvolvidos no âmbito da universidade

O Plano de gestão de dados é um documento que auxilia o pesquisador no planejamento e organização de sua pesquisa, como um todo, não somente no que diz respeito aos dados produzidos. Tendo em vista o exposto, a universidade poderia solicitar ao pesquisador de Sociologia ou de Química, a elaboração de um Plano de gestão de dados para quaisquer projetos realizados no âmbito da instituição. Esse plano nada mais é do que um documento administrativo complementar, no qual são genericamente descritos os principais procedimentos associados aos dados durante o ciclo da pesquisa científica.

O Plano de gestão de dados abrangeria informações relacionadas aos seguintes aspectos:

- tipo de dado que será coletado;
- descrição de conjuntos de dados e métodos para obtê-los;
- detalhamento de estratégias de preservação de dados;
- intencionalidade de permissão para o compartilhamento de dados produzidos, conforme cláusulas dispostas em licenças de dados.

Considerando as novas tecnologias que cada vez mais têm sido aplicadas no contexto de gestão acadêmica, o Plano de gestão de dados poderia ser submetido às universidades por meio de formulário eletrônico preenchido pelo pesquisador. Esse, por sua vez, necessitaria um sistema institucional para o envio do documento, de modo a centralizar todas as informações referentes ao projeto de pesquisa o qual pretende realizar na universidade. É importante dizer que o sistema desenvolvido para esse fim deveria apresentar plataforma amigável, dispor de tutoriais, ser acessível e sem ambiguidades para facilitar a sua utilização.

5.2 Prezar pela proteção ética de dados de pesquisa produzidos na instituição

Uma vez que se tem falado muito sobre acesso a dados abertos, deve-se ressaltar a importância da aplicação de mecanismos de proteção ética a determinados conjuntos de dados, tanto na área de Sociologia, quanto na área de Química.

Especificamente no campo da Sociologia, que se caracteriza, sobretudo, pela diversidade tipológica de dados que produz, é comum que pesquisadores trabalhem com dados que abordam temáticas sensíveis, os quais não seria ético serem expostos abertamente, sem nenhum tipo de tratamento técnico, como a anonimização de sujeitos. Esse processo é realizado com a finalidade de inviabilizar a identificação de indivíduos que consentiram participar de determinada pesquisa científica, incluindo a adoção de nomes fictícios para pessoas, instituições e até mesmo lugares. Isso é extremamente relevante no contexto de pesquisas relacionadas a tipologias criminais e execução penal, que, em geral, demandam algum nível de sigilo de informações antes da publicação de dados.

De modo semelhante, nem todos os dados contidos em diários de campo poderiam ser abertos sem uma verificação minuciosa dos conteúdos que apresentam, visto que tais diários podem ser entendidos como instrumentos de registro pessoalíssimo do pesquisador em relação aos fenômenos que estuda. Em outras palavras, trata-se da intimidade do pesquisador na qual sua visão de mundo é descrita sem filtros, devendo assim, não ser violada. Isso não quer dizer que dados produzidos em diários de campo não possam ser disponíveis abertamente, mas sim, que esses dados devam passar por tratamento técnico criterioso, com o propósito de minimizar quaisquer riscos de exposição indevida a pesquisadores e participantes da pesquisa.

De outra maneira, mas também relacionado ao aspecto da proteção ética de dados de pesquisa produzidos no contexto acadêmico, está a mitigação de vulnerabilidades na exposição de dados com caráter tecnológico, os quais não poderiam ser abertamente disponíveis de qualquer maneira. Isso porque trata-se de dados de uma natureza muito específica, em razão de dois aspectos principais: 1) são insumos para o pleiteio de titularização de patentes; 2) podem estar associados a negociações contratuais com instituições privadas para o desenvolvimento de novos produtos.

Sendo assim, dados com caráter tecnológico necessitam de tratamento diferenciado para não correrem o risco ético de uma comercialização antecipada por outros indivíduos ou empresas, ocasionando a perda de ineditismo, usurpação de ideias e exploração econômica indevida. Até porque, sabe-se que muitos países desenvolvidos possuem parque tecnológico muito mais avançado em determinadas áreas do que no Brasil, o que, provavelmente, contribuiria para o desenvolvimento de pesquisas e geração de dados de modo mais rápido, e com mais estrutura para a publicação de resultados científicos.

5.3 Oferecer suporte a pesquisadores para a construção de repositórios/bancos de dados, como alternativa para o compartilhamento e depósito de dados gerados em projetos de pesquisa

Repositórios disciplinares podem ser considerados locais seguros para o armazenamento de dados, além de oportunizarem o aumento na visibilidade de dados, promovendo assim, o compartilhamento dos mesmos na comunidade científica e o

reconhecimento de pesquisas conduzidas em diferentes áreas do conhecimento. Em vista do exposto, universidades deveriam incentivar o compartilhamento de dados de pesquisa em repositórios disciplinares sempre que possível. Porém, é sabido que nem todas as áreas dispõem desses tipos de plataformas. Sendo assim, caberia às universidades oferecer suporte a pesquisadores para a construção de repositórios/bancos de dados como alternativa para o compartilhamento e depósito de dados gerados em projetos de pesquisa.

Seria interessante que ao invés de construir repositórios de dados genéricos, as universidades estimulassem pesquisadores a conceber seus próprios repositórios de dados, ou a depositar dados produzidos em repositórios de suas comunidades. Entretanto, essas instituições deveriam dispor de condições estruturais e mecanismos tecnológicos adequados para viabilizar a agregação de dados disponibilizados nos repositórios disciplinares criados por pesquisadores.

Na Sociologia, são produzidos dados em diferentes formatos, extensões, em meio impresso ou digital. Trata-se de dados obtidos por meio de entrevistas, questionários, observações transcritas em diários de pesquisa, fotografias, vídeos, entre outros recursos. Na Química são coletados, principalmente, dados experimentais os quais costumam ser registrados em cadernos de laboratórios físicos ou digitais, para fins de documentação de processos científicos e comprovação da veracidade de informações obtidas. Logo, repositórios disciplinares devem ser robustos o suficiente para armazenar e veicular a diversidade de dados produzidos no contexto acadêmico, facilitando assim, a submissão de conjuntos de dados e posterior recuperação dos mesmos.

Desse modo, recomenda-se que dados sejam descritos a partir de metadados padronizados internacionalmente, sendo possível a adaptação de campos para melhor descrevê-los no contexto de cada área do conhecimento. Ademais, pode-se dizer que a descrição de dados associada à documentação sobre os mesmos em repositórios disciplinares contribui para que conjuntos de dados sejam facilmente localizados, analisados para fins de seleção, recuperados, tenham o seu histórico de procedimentos metodológicos conhecido e, em última análise, para que sejam reutilizados e impulsionem novos ciclos de produção científica.

5.4 Utilizar metabuscadores para recuperar e integrar conjuntos de dados e metadados depositados por pesquisadores em outros repositórios/bancos de dados

Repositórios que armazenam conjuntos de dados produzidos na área de Sociologia, Química, ou qualquer outra área, não devem veicular informações de forma isolada. Ao contrário, devem ser ponte entre diferentes plataformas para comunicar a produção científica e estabelecer relações interoperáveis com outros sistemas. Nesse sentido, universidades poderiam lançar mão de soluções tecnológicas ao utilizarem metabuscadores para recuperar e integrar conjuntos de dados e metadados depositados por seus pesquisadores em outros repositórios/bancos de dados (externos). Isso é relevante para a coleta de informações que podem ser insumos para tomadas de decisão sob diversos aspectos gerenciais da pesquisa, como a avaliação institucional a respeito de projetos científicos desenvolvidos.

5.5 Adotar estratégias de marketing para promover o reuso de dados abertos disponíveis em repositórios e dar visibilidade a pesquisas cujos dados encontram-se fechados, por meio da divulgação de seus metadados

A reutilização de dados de pesquisa é benéfica à comunidade científica, principalmente, por contribuir para a redução no tempo e custos da produção de pesquisas, otimizando recursos financeiros e evitando a duplicidades de estudos, além de contribuir para a comprovação de dados por meio da reprodução de procedimentos. Tanto na área de Sociologia, quanto na área de Química, observam-se tais benefícios. Entretanto, essa prática em ambas as disciplinas se demonstra timidamente, delimitando-se ao perímetro institucional do pesquisador ou dos grupos de pesquisa o qual integra. Isso evidencia a importância de o reuso de dados de pesquisa ser uma prática fomentada pelas universidades.

Nesse sentido, é imprescindível que universidades adotem estratégias de *marketing*, não somente para promover o reuso de conjuntos de dados abertos disponíveis em repositórios institucionais, mas também para dar visibilidade a pesquisas cujos dados encontram-se fechados, por meio da divulgação de seus metadados. Esses, por sua vez,

dão ênfase à pesquisa que está sendo realizada no âmbito das universidades, atribuindo a responsabilidade autoral a um pesquisador. Algumas dessas estratégias são elencadas da seguinte maneira:

- criação de serviço de disseminação seletiva de conjuntos de dados – serviço no qual membros da comunidade acadêmica se cadastrariam para receber metadados (em seu correio eletrônico ou aplicativo de mensagens de celular) sobre conjuntos de dados recém-submetidos a repositórios de dados de pesquisa, conforme as áreas de seu interesse;
- divulgação de repositórios de dados de pesquisa em *site* institucional, Intranet, mídias sociais, listas de discussões e por meio do envio de mensagens em aplicativos de celular.

5.6 Adotar estratégias de preservação de dados a longo prazo

Considerando resultados deste estudo nos quais verificou-se que a preservação de dados é, predominantemente, realizada de modo individual por pesquisadores da área de Sociologia e de Química, demonstra-se a importância de as universidades atentarem-se não somente à continuidade de acesso a resultados de estudos publicados, mas também aos dados produzidos ao longo da pesquisa científica.

Sendo assim, é fundamental que a universidade disponha de infraestrutura adequada para o armazenamento seguro dos dados produzidos na instituição. Isso porque, em geral, pesquisadores não possuem um Plano de preservação de dados a ser seguido e não dispõem de um repositório para armazenar os dados que produzem. Por conseguinte, muitos dados são perdidos em razão de danos ambientais (principalmente dados em meio impresso) e descontinuidade de leitores de mídias, ocasionada por sua obsolescência.

Outro aspecto relevante a esse contexto é que a preservação de dados deve contemplar não apenas dados abertos armazenados de forma segura em repositórios, mas também dados não abertos, os quais necessitam ser preservados, até mesmo para que possam ser disponibilizados abertamente, no futuro.

Considerando o exposto, é importante que universidades adotem estratégias de preservação dos dados produzidos em seu ambiente, dentre as quais pode-se destacar:

- institucionalizar a prática da preservação de dados de pesquisa, de modo a compartilhar responsabilidades entre a universidade e seus pesquisadores, não as incumbindo somente aos últimos;
- solicitar a pesquisadores a elaboração de um Plano de preservação de dados em projetos de pesquisa desenvolvidos no âmbito da universidade. Esse documento deve incluir procedimentos relacionados ao armazenamento seguro de dados, à realização de cópias de segurança e à ininterruptão das mídias que contém dados;
- providenciar local adequado para a guarda de documentos físicos, tais como arquivos setorizados;
- dispor de infraestrutura tecnológica adequada para a realização de cópias de segurança e migração de dados, quando necessário.

5.7 Viabilizar o reuso de dados abertos de pesquisa, prezando pelos direitos autorais sobre dados

Conforme observado nas áreas de Sociologia e de Química, a propriedade intelectual em relação a dados de pesquisa deve ser atribuída corretamente, sobretudo, no contexto da produção de dados referentes à inovação tecnológica para o desenvolvimento de novos produtos químicos. Portanto, o reuso de dados abertos de pesquisa deve ser promovido no ambiente acadêmico, porém, é importante que o mesmo ocorra em consonância com requisitos legais, geralmente contidos em licenças relacionadas à atribuição, uso e distribuição de dados.

5.8 Elaborar uma Política de gestão de dados abertos de pesquisa

Considerando a universidade como uma instituição que reúne diversos domínios disciplinares que, historicamente, se dedicam à geração do conhecimento científico, convém destacar a pluralidade de práticas referentes à produção, compartilhamento e uso de dados de pesquisa. Nesse contexto, ressalta-se a dificuldade na implementação de melhorias de processos, no tocante à gestão de dados abertos de pesquisa.

Com base no exposto, destaca-se a importância de universidades elaborarem um documento que formalize a Política de gestão de dados abertos a ser adotada em cada uma dessas instituições. Tal política seria o principal instrumento para nortear como pesquisadores e gestores institucionais devem proceder para viabilizar o acesso a dados de pesquisa, bem como a preservação dos mesmos.

É interessante pensar que a falta de uma Política de gestão de dados abertos de pesquisa em uma universidade contribui para uma série de aspectos negativos, tais como o baixo compartilhamento de dados que poderiam desencadear novos ciclos da produção científica, o aumento de custos na pesquisa, a duplicidade de trabalhos, a diminuição na visibilidade de publicações científicas, a indisponibilidade de dados para a checagem da veracidade de informações veiculadas, a perda de dados em razão da ausência de um Plano de preservação de dados a longo prazo, e a falta de discussões sobre o tópico da abertura de dados científicos entre pesquisadores e gestores institucionais.

Acrescente-se, oportunamente, que uma Política de gestão de dados abertos no contexto de uma universidade deve ser genericamente construída para abranger a pluralidade das áreas do conhecimento, e suficientemente específica para incluir os aspectos mais relevantes de cada uma delas em relação aos dados.

5.9 Implantar serviço de gestão de dados de pesquisa

Nota-se que muitos pesquisadores, tanto de Sociologia, quanto de Química, trabalham individualmente com os dados gerados durante o ciclo da pesquisa, sendo os principais responsáveis pelo planejamento do estudo, o compartilhamento de dados em repositórios e pela adoção de estratégias a fim de preservá-los.

Considerando as demandas institucionais crescentes de cientistas no âmbito administrativo, do ensino, da pesquisa e extensão, a gestão de dados de pesquisa deveria ser compartilhada entre universidades e pesquisadores, de modo a garantir o acesso e preservação de dados a longo prazo. Sendo assim, universidades poderiam implantar um serviço de gestão de dados de pesquisa, integrado por profissionais capacitados de diferentes setores, tais como bibliotecários e analistas de sistemas.

Dentre as atividades que poderiam ser desempenhadas pelo serviço de gestão de dados de pesquisa, destacam-se as seguintes:

- auxílio técnico a pesquisadores na elaboração do Plano de gestão de dados de pesquisa;
- processamento técnico de dados de pesquisa, o que inclui procedimentos como a digitalização de documentos, transcrição de entrevistas, organização de conteúdos de vídeo e imagem, anonimização de indivíduos, verificação da necessidade de sigilo em alguns dados com caráter de inovação tecnológica, atribuição de metadados para descrever e documentar conjuntos de dados armazenados em repositórios de dados, entre outros;
- auxílio técnico a pesquisadores na elaboração do Plano de preservação de dados de pesquisa;
- realização de cópias de segurança e migração de dados, visando preservá-los a longo prazo;
- orientação quanto aos direitos de propriedade intelectual sobre dados de pesquisa;
- divulgação à comunidade acadêmica de conjuntos de dados abertos de pesquisa recém-disponibilizados nas coleções do repositório de dados;
- orientação geral a pesquisadores sobre a gestão de dados de pesquisa, informando-os sobre as melhores práticas realizadas nesse sentido.

5.10 Capacitar equipes de bibliotecários para a atuação em serviços de gestão de dados de pesquisa

Considerando o bibliotecário como o profissional que trabalha para a disseminação do conhecimento por meio da disponibilização de fontes de informação, as quais incluem conjuntos de dados de pesquisa, observa-se que é cada vez mais importante a integração de serviços tradicionais àqueles relacionados à gestão de dados de pesquisa.

Nessa direção, dados de pesquisa podem ser vistos como qualquer outro registro do conhecimento humano disponível que necessita ser mediado por um bibliotecário para tornar-se acessível a usuários que possuem uma necessidade de informação específica. Para tanto, é fundamental que tal profissional se sinta participante desse processo,

contribuindo assim, para que dados recebam tratamento técnico adequado, mantendo-se acessíveis e preservados a longo prazo. Trata-se de criar as condições necessárias para que os mesmos estejam amplamente disponíveis para o reuso, e, por conseguinte, ser insumos para a geração de novos ciclos de produção científica.

Para isso, é importante que bibliotecários estejam aptos para atender às demandas decorrentes da implantação de um serviço de gestão de dados de pesquisa no ambiente acadêmico, principalmente no que se refere à disponibilização de dados abertos de pesquisa. Tal aptidão pode ser alcançada, principalmente, por meio de treinamentos institucionais, ciclos de palestras, rodas de conversa e encontros nacionais de bibliotecários que abordem aspectos como Ciência aberta, abertura de dados e gestão de dados de pesquisa.

5.11 Fomentar discussões sobre a gestão de dados abertos de pesquisas na comunidade acadêmica

Apesar de o tópico da Ciência aberta aparecer em discussões originárias do convívio acadêmico entre pesquisadores da Sociologia, e de ser mencionado em alguns eventos na área da Química, nota-se que esse não é um tema recorrentemente pautado em ambas as disciplinas. De modo geral, observa-se que a importância da abertura de dados de pesquisa e da realização de uma efetiva gestão sobre esses dados não têm sido amplamente abordadas no contexto universitário.

Sendo assim, algumas dúvidas relacionadas à gestão de dados abertos persistem entre pesquisadores, tais como: há algum repositório para o armazenamento de dados? Quais as melhores estratégias adotadas para a sua preservação? Como compartilhá-los de forma ética? Como reutilizá-los sem ferir direitos de propriedades sobre dados?

É nesse contexto de incertezas, que se recomenda que universidades promovam a discussão sobre a importância da gestão de dados abertos de pesquisa dentro da comunidade acadêmica. É preciso refletir o que de fato significa o fenômeno da Ciência aberta e quais os seus impactos na produção e compartilhamento do conhecimento científico a longo prazo. Ademais, é fundamental analisar como serviços de gestão de dados poderão ser desempenhados em universidades, quais recursos serão necessários e

qual infraestrutura tecnológica será a mais adequada para abranger a diversidade tipológica de dados que são produzidos.

O fomento de discussões sobre a gestão de dados abertos de pesquisa em universidades pode ser realizado de várias maneiras, dentre as quais pode-se destacar:

- realização de eventos como palestras, oficinas e rodas de conversa sobre o tema, voltados à comunidade acadêmica;
- treinamentos a pesquisadores sobre o depósito de dados abertos em repositórios de dados;
- treinamentos a pesquisadores sobre a utilização de repositórios de dados abertos de pesquisa;
- ministração de disciplinas de módulo livre associadas à grade curricular dos cursos de graduação e pós-graduação.

5.12 Implantar sistema de recompensas para a valorização do compartilhamento de dados abertos de pesquisa

Considerando que pesquisadores de Sociologia e de Química sugeriram a implantação de um sistema de recompensas como mecanismo de incentivo para o compartilhamento de dados abertos de pesquisa, aqui se propõe que os cientistas que optarem pela abertura dos conjuntos de dados em repositórios sejam beneficiados de alguma maneira, como a progressão funcional e o aumento nos valores de recursos financeiros para o desenvolvimento de projetos de pesquisa.

Cabe ressaltar, ainda, que não seria interessante punir aqueles pesquisadores que se opusessem, de algum modo, a compartilhar abertamente os dados que produzem. Todavia, esses ficariam em desvantagem em relação a outros pesquisadores que disponibilizassem seus dados em algum repositório, quando ambos concorressem a editais internos de financiamentos a projetos de pesquisa.

5.13 Gerir recursos não somente para a divulgação de resultados de pesquisa publicados, mas também de dados de pesquisa produzidos na instituição

Assim como fontes de informação que registram resultados de pesquisa publicados, dados de pesquisa demandam investimento de recursos financeiros, tecnológicos, de pessoal e determinada logística para a sua disponibilização. Sejam dados produzidos no contexto da Sociologia, Química, ou quaisquer outras áreas, é importante ter uma visão sistêmica da comunicação científica para que a gestão de recursos institucionais não se concentre apenas em resultados de pesquisa publicados, mas que inclua o aspecto dos dados de pesquisa. Até porque, em geral, dados são construídos ao longo de projetos de pesquisa, sendo ponte para a obtenção de resultados os quais costumam ser publicados posteriormente.

5.14 Gerir dados abertos de pesquisa de modo a considerar diferenças disciplinares

A necessidade de discussão sobre o tema da Ciência aberta e, especificamente, dos dados abertos de pesquisa, foi evidenciada neste estudo, tanto na área de Sociologia, quanto na área de Química. Entretanto, o contexto universitário não diz respeito a apenas duas áreas, mas sim a várias áreas que possuem perspectivas distintas para entender os fenômenos que circundam a humanidade ao longo do tempo. Conhecer intimamente tais perspectivas, no sentido de como dados são produzidos, compartilhados e utilizados, pode auxiliar a prática de uma gestão de dados abertos de pesquisa muito mais assertiva, do que não as considerar.

5.15 Promover o depósito de dados abertos de pesquisa em repositórios, por meio de vias dialógicas

Apesar de alguns pesquisadores de Sociologia e Química argumentarem que dados produzidos com recursos públicos deveriam, necessariamente, estar abertos à ampla audiência, observa-se em ambas as disciplinas que um dos caminhos mais proveitosos para

promover o depósito de dados abertos de pesquisa em repositórios de dados é o diálogo entre universidade e pesquisador.

É interessante notar que reflexões sobre os potenciais efeitos da abertura de dados de pesquisa e seus principais benefícios são vias mais receptíveis aos pesquisadores de Sociologia e de Química, do que decisões obrigatórias. Em geral, o termo “obrigatório” está associado a uma imposição resultante de uma escolha não democrática, ou não participativa. Logo, sugere-se que no contexto de ambas as áreas, o depósito de dados abertos de pesquisa em repositórios seja promovido por meio de diálogo com cientistas, e não por obrigação.

6 CONCLUSÕES

O presente estudo teve como objetivo geral propor diretrizes para políticas institucionais de gestão de dados abertos de pesquisa em universidades. De modo a atingi-lo, foram definidos três objetivos específicos: 1) identificar práticas de gestão de dados de pesquisa realizadas em diferentes áreas, com base na literatura científica; 2) analisar práticas de gestão de dados de pesquisa realizadas por pesquisadores em diferentes áreas; 3) descrever fatores que influenciam a abertura de dados de pesquisa, segundo pesquisadores em diferentes áreas. As conclusões da pesquisa referentes a cada um dos objetivos mencionados serão apresentadas nos tópicos seguintes:

6.1 Identificar práticas de gestão de dados de pesquisa realizadas em diferentes áreas, com base na literatura científica

O objetivo foi alcançado por meio de revisão sistemática da literatura realizada a partir da análise de 14 artigos indexados na base de dados *Library and Information Science Abstracts* (LISA), publicados entre 2015 e 2020. Sendo assim, foram identificadas doze práticas de gestão de dados de pesquisa as quais podem ser descritas, resumidamente, da seguinte maneira:

- **planejamento** – atividades relacionadas ao planejamento da gestão de dados e de resultados de pesquisas;
- **coleta** – procedimentos de gestão de dados associados à produção/obtenção de dados;
- **processamento técnico** – tratamento técnico de dados de pesquisa. Inclui procedimentos de inclusão de metadados, anonimização de indivíduos e tradução de dados;
- **compartilhamento de dados** – procedimentos e plataforma que viabilizam o compartilhamento de dados de pesquisa, de modo adequado;
- **preservação de dados** – atividades que promovem a preservação de dados de pesquisa a longo prazo;

- **descoberta de dados** – uso de estratégias que contribuem para o aumento da visibilidade de conjuntos de dados de pesquisa;
- **reuso de dados** – adoção de estratégias que potencializam o reuso de dados de pesquisa;
- **monitoramento e controle da gestão de dados de pesquisa** – avaliação e monitoramento de atividades relacionadas à gestão de dados de pesquisa;
- **políticas para a gestão de dados de pesquisa** – elaboração de documentos que orientem como a gestão de dados deve ser conduzida;
- **suporte à análise de dados** – apoio institucional direcionado a pesquisadores no processo de análise de dados;
- **treinamento oferecido a profissionais da informação** – capacitação de profissionais da informação sobre o tópico da gestão de dados de pesquisa;
- **treinamento oferecido a pesquisadores** – capacitação de pesquisadores sobre o tópico da gestão de dados de pesquisa;

6.2 Analisar práticas de gestão de dados de pesquisa realizadas por pesquisadores em diferentes áreas

O objetivo foi alcançado, podendo ser averiguado a partir dos resultados obtidos por meio das entrevistas realizadas com bolsistas de produtividade em pesquisa - (nível PQ-2) do CNPq das áreas de Química e de Sociologia vinculados a universidades federais brasileiras. Utilizando o método da teoria fundamentada e seus processos – codificação aberta, codificação axial, codificação seletiva e elaboração de teoria substantiva – foi possível concluir os seguintes aspectos em relação às práticas de gestão de dados de pesquisa realizadas em ambas as áreas:

Em Sociologia:

- a ética na pesquisa é o elemento central para a compreensão das práticas de gestão de dados de pesquisa realizadas na área;
- a Ciência aberta é um tópico que vem sendo discutido no convívio de pesquisadores da área, principalmente em grupos de pesquisa, departamentos acadêmicos e

associações científicas. Nesse contexto, cientistas procuram entender quais as reais implicações da Ciência aberta em relação aos aspectos éticos na pesquisa;

- apesar de haver uma predominância de instituições públicas no fomento à pesquisa na área de Sociologia, alguns pesquisadores são solicitados a entregar um Plano de gestão de dados, comunicando quais estratégias éticas serão utilizadas em relação a dados que serão produzidos. Todavia, tal demanda ainda não é uma realidade amplamente difundida na área;
- a produção de dados no campo da Sociologia caracteriza-se, sobretudo, por sua diversidade tipológica. Nesse contexto, muitos dados tratam de temáticas sensíveis, não podendo ser compartilhados abertamente sem a observação de alguma prescrição ética;
- dados de temáticas sensíveis, em geral, necessitam de maior observância a critérios éticos, porém, não costumam ser descritos por pesquisadores conforme padrões internacionais de metadados específicos para a área de Sociologia. Cada pesquisador utiliza o seu padrão para a descrição dos dados que produz;
- a documentação sobre dados gerados em pesquisas na área viabilizam a consulta ao histórico de procedimentos realizados ao longo do trabalho científico, facilitando assim, a verificação de inconsistências metodológicas e éticas em pesquisas;
- o compartilhamento de dados na área de Sociologia é predominantemente restrito ao contexto de grupos de pesquisa, dando-se por meio de canais informais. É comum que pesquisadores estabeleçam relações de confiança com quem desejam compartilhar os dados que produzem, avaliando também a reputação ética de outros cientistas;
- a preservação de dados inclui aspectos éticos associados a padrões de segurança de dados, especialmente no contexto daqueles que tratam sobre temáticas sensíveis. Porém, observa-se que pesquisadores na área de Sociologia lançam mão de técnicas individuais para esse fim. Em geral, não há diretrizes dispostas em políticas institucionais de preservação de dados, bem como não é comum a exigência de um Plano de preservação de dados a cientistas;
- o reuso de dados de pesquisa no campo da Sociologia ocorre predominante por meio de colaboração científica. Em geral, para que dados produzidos com apoio

financeiro de alguma agência de fomento possam ser reutilizados, acordos éticos exigem a citação da instituição para publicações derivadas;

- considerando as diferenças disciplinares como um fio condutor para a gestão de dados abertos de pesquisa em universidades, menciona-se que no campo da Sociologia é dispensada muita atenção a aspectos éticos por parte de pesquisadores em relação à anonimização de dados e participantes de estudos.

Em Química:

- a comprovação de dados é o elemento central para o entendimento das práticas de gestão de dados de pesquisa realizadas na área;
- restrições como a exigência do pagamento para a publicação de dados abertos dificultam o reuso dos mesmos, e, por conseguinte, a disponibilidade de fontes confiáveis que poderiam ser utilizadas na comprovação de estudos realizados;
- considerando que o planejamento sobre dados de pesquisa na área geralmente está associado ao recebimento de recursos financeiros provenientes de agências de fomento para a execução de pesquisas, muitas vezes, esses dados assumem um caráter comprobatório diante de tais instituições;
- cadernos de laboratório ocupam um lugar de relevância no processo de coleta de dados, uma vez que neles são registrados os principais procedimentos realizados em um pesquisa, o que é igualmente importante para a comprovação de resultados obtidos;
- em geral, pesquisadores na área da Química adotam procedimentos próprios para a descrição dos dados que produzem, sem o uso de metadados os quais poderiam facilitar a localização, análise e posterior comprovação de resultados de pesquisas;
- geralmente, a documentação de procedimentos relacionados aos dados de pesquisa na área de Química está associada à verificação de dados para fins comprobatórios;
- repositórios de dados podem contribuir para o aumento na visibilidade de pesquisas, sendo mais um canal que viabiliza a recuperação, análise e comprovação de informações que veiculam;
- em geral, pesquisadores na área da Química adotam procedimentos próprios para assegurar a preservação dos dados que geram, acarretando riscos que comprometem o reuso dos mesmos a longo prazo. Por conseguinte, não há ampla

segurança de que dados continuarão disponíveis para consulta, análise e posterior verificação, a fim de que sejam comprovados;

- muitos pesquisadores na área da Química reutilizam informações disponibilizadas em trabalhos publicados na literatura. O reuso de dados brutos não é uma atividade habitual;
- considerando a importância das diferenças disciplinares no âmbito da gestão de dados abertos de pesquisa em universidades, destaca-se que no campo da Química dá-se muita atenção ao aspecto da propriedade intelectual sobre dados, em especial, no que se refere a elementos que comprovem a execução de uma pesquisa e aos termos legais que condicionam o seu reuso em outros trabalhos.

6.2.1 Aspectos convergentes e divergentes em relação às práticas de gestão de dados de pesquisa realizadas por pesquisadores na área de Sociologia e de Química

Os resultados obtidos em relação às práticas de gestão de dados de pesquisa realizadas por pesquisadores na área de Sociologia e de Química demonstram convergências em alguns aspectos, os quais podem ser destacados os seguintes:

- o planejamento relacionado à produção de pesquisas em ambas as áreas concentra-se na descrição de processos administrativos, relacionados, principalmente, à distribuição de recursos financeiros que foram recebidos. A exigência de um Plano de gestão de dados a pesquisadores ainda não é uma realidade amplamente conhecida no contexto de universidades;
- geralmente, a descrição de dados nas duas áreas é realizada pelo próprio pesquisador que os gerou, não sendo habitual o uso de metadados padronizados para esse fim;
- estratégias de preservação de dados são predominantemente individuais (do próprio pesquisador que gerou determinado conjunto de dados), não sendo práticas institucionalizadas;
- é muito importante que a gestão de dados de pesquisa em universidades estruture-se no argumento da pluralidade disciplinar, de modo a considerar diferenças na

produção, compartilhamento e uso de dados, assim como foi observado neste trabalho no contexto das áreas de Sociologia e Química.

De outro modo, os resultados alcançados referentes às práticas de gestão de dados de pesquisa realizadas por pesquisadores na área de Sociologia e de Química apresentam divergências em alguns pontos, os quais podem ser ressaltados da seguinte maneira:

- a natureza, os formatos, os tipos e outras características dos dados produzidos no campo da Sociologia diferem-se dos dados gerados no campo da Química. Por isso, é imprescindível que repositórios institucionais disponham de infraestrutura suficiente para viabilizar o armazenamento, descrição e disponibilização de dados produzidos em diversas áreas;
- pesquisadores entrevistados na área de Química demonstraram ter mais experiências no compartilhamento de dados abertos em repositórios/plataformas institucionais do que pesquisadores no campo da Sociologia, desde que os dados compartilhados não tratem de inovação tecnológica.

6.3 Descrever fatores que influenciam a abertura de dados de pesquisa, segundo pesquisadores em diferentes áreas

O objetivo foi atingido, podendo ser verificado a partir dos resultados obtidos por meio das entrevistas concedidas por bolsistas de produtividade em pesquisa - (nível PQ-2) do CNPq das áreas de Química e de Sociologia vinculados a universidades federais brasileiras. Lançando mão do método da teoria fundamentada e seus processos – codificação aberta, codificação axial, codificação seletiva e elaboração de teoria substantiva – pode-se concluir os seguintes aspectos em relação aos fatores que influenciam a abertura de dados de pesquisa em ambas as áreas:

Em Sociologia:

- os fatores que estimulam a abertura de dados de pesquisa são: possibilidade de aprimoramento de estudos na área; recebimento de financiamento público da pesquisa; liderança institucional para conduzir os processos relacionados à abertura de dados;

- os aspectos favoráveis à obrigatoriedade da abertura de dados a pesquisadores estabelecida por instituições acadêmicas e científicas são: a possibilidade de comprovação de dados por editores científicos e o argumento de que pesquisas fomentadas com dinheiro público deveriam ter os dados disponíveis abertamente;
- os fatores que inibem a abertura de dados de pesquisa são: existência de riscos éticos; falta de infraestrutura para esse fim; o tópico da abertura de dados ainda ser pouco explorado na área; exigências de pagamento para publicações em acesso aberto;
- o aspecto desfavorável à obrigatoriedade da abertura de dados a pesquisadores estabelecida por instituições acadêmicas e científicas é: o argumento o qual o pesquisador tem livre-arbítrio para decidir a respeito da abertura dos dados que produz;
- a infraestrutura para a disponibilização de dados abertos é o elemento central para o entendimento dos fatores que influenciam a abertura de dados de pesquisa no contexto da área;
- é imprescindível que cientistas tenham à sua disposição a infraestrutura apropriada para viabilizar a abertura de dados de pesquisa, tornando-os reutilizáveis para fim de aprimoramento de estudos na área e comprovação de resultados;
- a infraestrutura para a abertura de dados de pesquisa é fundamental para assegurar a proteção de dados, no que se refere a aspectos éticos;
- em geral, cientistas da área de Sociologia não têm a seu dispor recursos suficientes para a execução de suas pesquisas, o que reflete na falta de infraestrutura adequada para a abertura de dados;
- observa-se uma contínua redução de recursos previstos em orçamentos públicos destinados ao campo da educação e desenvolvimento científico-tecnológico, o que contribui para a falta de estrutura relacionada à produção e disponibilização de dados abertos de pesquisa;
- observa-se a falta de uma liderança institucional na área de Sociologia para orientar a gestão de dados de pesquisa no contexto brasileiro, evidenciando assim, a necessidade da realização de treinamentos sobre o tema a

pesquisadores e profissionais da informação;

- mesmo havendo infraestrutura adequada, a obrigatoriedade de disponibilização de dados abertos demandada por instituições acadêmicas e científicas a pesquisadores encontra resistência no campo da Sociologia. Trata-se do apreço da área por decisões democraticamente construídas.

Em Química:

- os fatores que estimulam a abertura de dados de pesquisa são: possibilidade de aprimoramento de estudos na área; liderança institucional para conduzir os processos relacionados à abertura de dados; aumento na visibilidade de pesquisas; disponibilidade de recursos destinados à abertura de dados;
- os aspectos favoráveis à obrigatoriedade da abertura de dados a pesquisadores estabelecida por instituições acadêmicas e científicas são: o argumento de que pesquisas fomentadas com dinheiro público deveriam ter os dados disponíveis abertamente, e a possibilidade de comprovação de dados por editores científicos;
- os fatores que inibem a abertura de dados de pesquisa são: concessão de dados de natureza tecnológica; existência de riscos éticos; custos da disponibilização de dados abertos de pesquisa;
- o aspecto desfavorável à obrigatoriedade da abertura de dados a pesquisadores estabelecida por instituições acadêmicas e científicas é: o argumento de que a abertura de dados deve ser uma concessão voluntária do pesquisador que os produziu;
- vulnerabilidades na exposição de dados com caráter de inovação tecnológica é o elemento central para a compreensão dos fatores que influenciam a abertura de dados de pesquisa no contexto da área;
- para que dados sejam disponibilizados abertamente, possibilitando seu reuso para o aprimoramento de pesquisas na área de Química, comprovação de informações junto a editores científicos e para o aumento na visibilidade de pesquisas, é fundamental verificar se existem vulnerabilidades na exposição de dados com caráter tecnológico, os quais não poderiam ser compartilhados de qualquer modo;
- muitos dados com caráter de inovação tecnológica serão insumos para a solicitação de patenteamento de produtos, não podendo ser amplamente

disponibilizados por determinado período. Sendo assim, tais dados necessitam de proteção para minimizar riscos éticos relacionados a uma publicação antecipada dos mesmos;

- a disponibilização de dados abertos de pesquisa envolve custos para a realização de atividades como o processamento técnico de dados com caráter tecnológico, com a finalidade de minimizar vulnerabilidades em sua exposição;
- reduções orçamentárias impactam negativamente a produção de pesquisas brasileiras e, por conseguinte, a disponibilização de dados abertos na área;
- o fortalecimento de discussões sobre a abertura de dados por parte de uma liderança institucional contribuiria para mitigar vulnerabilidades relacionadas a esse processo;
- cientistas na área de Química não estão de acordo com uma possível obrigatoriedade de disponibilização de dados demandada por instituições científicas e acadêmicas. Para muitos pesquisadores nesse campo, trata-se de uma imposição que não seria bem recepcionada pelos mesmos.

6.3.1 Aspectos convergentes e divergentes em relação aos fatores que influenciam a abertura de dados de pesquisa na área de Sociologia e de Química

Os resultados alcançados em relação aos fatores que influenciam a abertura de dados de pesquisa na área de Sociologia e de Química demonstram convergências em alguns aspectos, os quais podem ser destacados os seguintes:

- as falas de pesquisadores entrevistados demonstram que a Ciência aberta (especialmente no tocante à abertura de dados) é um campo bastante desafiador. Apesar de o estudo ter considerado perspectivas nas duas áreas analisadas, destaca-se a existência de especificidades na produção e compartilhamento de dados, não apenas relacionadas a domínios disciplinares, mas também a linhas e projetos de pesquisa desempenhados por pesquisadores, principalmente no contexto de universidades situadas no sul-global;
- é claramente observada a importância em se discutir a abertura de dados de pesquisa no âmbito da comunidade acadêmica, em termos de benefícios,

contradições, disponibilidade de recursos e opiniões de pesquisadores e gestores da informação;

- reduções orçamentárias no campo da educação e ciência afetam negativamente a disponibilização de infraestruturas adequadas que viabilizam a abertura de dados de pesquisa;

A ênfase na propriedade intelectual sobre dados e na possibilidade de solicitação de patentes por parte de pesquisadores de Química são alguns dos principais aspectos divergentes em relação à área de Sociologia. Além disso, trata-se da diferença originária e tipológica entre dados experimentais produzidos, geralmente, em ambientes de laboratório e dados caracterizados, sobretudo, pela diversidade de métodos utilizados para a sua obtenção, como ocorre no contexto da Sociologia.

6.4 Diretrizes que podem vir a ser instituídas para promover a abertura de dados de pesquisa em diferentes áreas do conhecimento

O estudo pretendeu buscar respostas à pergunta: Quais diretrizes podem vir a ser instituídas para promover a abertura de dados de pesquisa em diferentes áreas do conhecimento? A partir dos resultados obtidos neste trabalho em relação às práticas de gestão de dados de pesquisa e aos fatores que influenciam a abertura de dados no campo da Sociologia e da Química, foi possível responder à pergunta referida, a partir da elaboração de um conjunto de diretrizes para políticas institucionais de gestão de dados abertos de pesquisa no contexto de universidades. Essas diretrizes são:

- 1) solicitar ao pesquisador a elaboração de um Plano de gestão de dados em projetos de pesquisa desenvolvidos no âmbito da universidade;
- 2) prezar pela proteção ética de dados de pesquisa produzidos na instituição;
- 3) oferecer suporte a pesquisadores para a construção de repositórios/bancos de dados, como alternativa para o compartilhamento e depósito de dados gerados em projetos de pesquisa;
- 4) utilizar metabuscadores para recuperar e integrar conjuntos de dados e metadados depositados por pesquisadores em outros repositórios/bancos de dados;

- 5) adotar estratégias de *marketing* para promover o reuso de dados abertos disponíveis em repositórios e dar visibilidade a pesquisas cujos dados encontram-se fechados, por meio da divulgação de seus metadados;
- 6) adotar estratégias de preservação de dados a longo prazo;
- 7) viabilizar o reuso de dados abertos de pesquisa, prezando pelos direitos autorais sobre dados;
- 8) elaborar uma Política de gestão de dados abertos de pesquisa;
- 9) implantar serviço de gestão de dados de pesquisa;
- 10) capacitar equipes de bibliotecários para a atuação em serviços de gestão de dados de pesquisa;
- 11) fomentar discussões sobre a gestão de dados abertos de pesquisas na comunidade acadêmica;
- 12) implantar sistema de recompensas para a valorização do compartilhamento de dados abertos de pesquisa;
- 13) gerir recursos não somente para a divulgação de resultados de pesquisa publicados, mas também de dados de pesquisa produzidos na instituição;
- 14) gerir dados abertos de pesquisa de modo a considerar diferenças disciplinares;
- 15) promover o depósito de dados abertos de pesquisa em repositórios, por meio de vias dialógicas.

6.5 Sugestões de estudos futuros

Como se pode observar, o estudo atingiu o objetivo geral de “propor diretrizes para políticas institucionais de gestão de dados abertos de pesquisa em universidades”. Entretanto, convém dizer que generalizações das conclusões obtidas na pesquisa podem não ter os mesmos efeitos e impactos em disciplinas de outras áreas. Por isso, destaca-se que o presente trabalho não se esgota aqui. Ao contrário, abre margem para a realização de outros estudos associados ao tema abordado. Como sugestão para futuros estudos, recomenda-se:

- investigar como o fenômeno da Ciência aberta tem sido discutido na literatura científica ao longo do tempo e em diferentes países, a partir de comparações

entre os mesmos;

- caracterizar a produção de dados de pesquisa em diferentes áreas;
- identificar padrões e requisitos de metadados para a descrição de dados de pesquisa em diferentes áreas;
- analisar processos de armazenamento, compartilhamento e recuperação de dados em repositórios institucionais de dados abertos de pesquisa;
- avaliar o impacto de citações a conjuntos de dados abertos disponíveis em repositórios institucionais ou periódicos de dados;
- descrever estratégias de preservação de dados de pesquisa a longo prazo;
- investigar estratégias para a promoção do reuso de dados de pesquisa;
- analisar relações entre direitos autorais sobre dados de pesquisa e acesso aberto;
- analisar políticas para a gestão de dados de pesquisa, comparando-as em relação às diretrizes que determinam;
- investigar a gestão de dados de pesquisa em universidades, sob a perspectiva de gestores da informação e do conhecimento.

REFERÊNCIAS

ABDULDAYAN, F. J. *et al.* Research data management practices of chemistry researchers in federal universities of technology in Nigeria. **Digital Library Perspectives**, v. 37, n. 4, p. 328-348, 2021.

AGÊNCIA PARA A MODERNIZAÇÃO ADMINISTRATIVA (AMA). **Guia de dados abertos**. 2016. 14 p. Disponível em: https://www.ama.gov.pt/documents/24077/24804/guia_introdu__o_dados_abertos_ama.pdf/9b40b98c-4935-471b-af5d-f6f6a6556edc0. Acesso em: 03 set. 2020.

ALCALÁ, M; ANGLADA, L. FAIR x FAIR. Una aproximación pragmática a la gestión de datos de investigación en modo FAIR. **Anuario ThinkEPI**, v. 13, p. 1-4, sep. 2019. Disponível em: <https://dialnet.unirioja.es/servlet/articulo?codigo=7182733>. Acesso em: 17 jun. 2021.

ALONSO ARÉVALO, J.; SUBIRATS COLL, I.; MARTÍNEZ CONDE, M. L. **Informe APEI sobre acceso abierto**. Espanha: Asociación Profesional de Especialistas en Información, 2008. 63 p. Disponível em: <http://eprints.rclis.org/12507/1/informeapeiaccesoabierto.pdf>. Acesso em: 27 jan. 2021.

ANALYTICS CLARIVATE. **Research in Brazil: Funding excellence: Analysis prepared on behalf of CAPES by the Web of Science Group**. [s.l.]: Web of Science Group, 2019. 41 p.

APPOLINÁRIO, F. **Dicionário de metodologia científica: um guia para a produção do conhecimento científico**. 2. ed. São Paulo: Atlas, 2011.

AVUGLAH, B. K.; UNDERWOOD, P. G. Research data management (RDM) capabilities at the University of Ghana, Legon. **Library Philosophy and Practice**, p. 1-18, apr. 2019. Disponível em: https://repository.up.ac.za/bitstream/handle/2263/71138/Avuglah_Research_2018.pdf?sequence=1. Acesso em: 17 jun. 2021.

AYDINOGLU, A. U.; DOGAN, G.; TASKIN, Z. Research data management in Turkey: perceptions and practices. **Library Hi Tech**, v. 35, n. 2, p. 271-289. 2017. Disponível em: http://www.openaccess.hacettepe.edu.tr:8080/xmlui/bitstream/handle/11655/10805/10-1108_LHT-11-2016-0134.pdf?sequence=1. Acesso em: 17 jun. 2021.

BALL, A. **How to license research data**. [s.l.]: Digital Curation Centre, 2014. 15 p. Disponível em: <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.220.6839&rep=rep1&type=pdf>. Acesso em: 16 jun. 2021.

BANERJEE, I; BABINI, D.; AGUADO, E. Tesis a favor de la consolidación del Acceso Abierto como una alternativa de democratización de la ciencia en América Latina. In: SUBER, P. **Acceso Abierto**. Traducción de Remedios Melero. Toluca, Estado de México: Universidad Autónoma del Estado de México, 2015. p. 13-48 (Colección Cuadernos Institucionales). Disponível em: <https://digital.csic.es/bitstream/10261/121428/1/ACCESO%20ABIERTO.pdf>. Acesso em: 27 jan. 2021.

BARBROW, S.; BRUSH, D.; GOLDMAN, J. Research data management and services: resources for novice data librarians. **College & Research Libraries News**, v. 78, n. 5, p. 274-278, may 2017. Disponível em: <https://crln.acrl.org/index.php/crlnews/article/view/16660/18115>. Acesso em: 16 fev. 2021.

BELLEGARD, M. I. ERDMAS: An exemplar-driven institutional research data management and analysis strategy. **International Journal of Information Management**, v. 50, p. 337-340, 2020. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0268401219307327>. Acesso em: 17 jun. 2021.

BERTIN, P. R. B.; VISOLI, M. C.; DRUCKER, D. P. A gestão de dados de pesquisa no contexto da e-science: benefícios, desafios e oportunidades para organizações de P&D. **Ponto de Acesso**, v. 11, n. 2, p. 34-48, 2017. Disponível em: <https://portalseer.ufba.br/index.php/revistaici/article/view/21449>. Acesso em: 03 set. 2020.

BETHESDA STATEMENT ON OPEN ACCESS PUBLISHING. **Meeting on Open Access Publishing**, 2003. Disponível em: <http://legacy.earlham.edu/~peters/fos/bethesda.ht>. Acesso em: 27 jan. 2021.

BHATTACHERJEE, A. **Social science research: principles, methods, and practices**. USF Tampa Bay Open Access Textbooks Collection. Book 3. 2012.

BONI, V.; QUARESMA, S. J. Aprendendo a entrevistar: como fazer entrevistas em Ciências Sociais. **Revista Eletrônica dos Pós-Graduandos em Sociologia Política da UFSC**, v. 2, n. 1, p. 68-80, jan./jul. 2005. Disponível em: <https://periodicos.ufsc.br/index.php/emtese/article/viewFile/18027/16976+&cd=1&hl=pt-BR&ct=clnk&gl=br>. Acesso em: 10 set. 2020.

BOREN, S. A.; MOXLEY, D. Systematically reviewing the literature: building the evidence for health care quality. **Missouri medicine**, v. 112, n. 1, p. 58-62, 2015. Disponível em: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6170102/pdf/ms112_p0058.pdf. Acesso em: 03 set. 2020.

BORGMAN, C. L. **Big data, little data, no data: scholarship in the networked world**. Cambridge, Massachusetts: The MIT Press, 2015. 416 p.

BRASIL. Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). Resolução Normativa nº 28, de 18 de dezembro de 2015. Resolve estabelecer as normas gerais e específicas para as modalidades de bolsas individuais no País relacionadas no Anexo I. **Diário Oficial da União**, Brasília, DF, 21 dez. 2015. Seção 1, p. 45. Disponível em: <https://pesquisa.in.gov.br/imprensa/jsp/visualiza/index.jsp?data=21/12/2015&jornal=1&pagina=45&totalArquivos=240>. Acesso em: 03 set. 2020.

BRASIL. Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). **Bolsas em Curso**. Disponível em: http://plsq11.cnpq.br/divulg/RESULTADO_PQ_102003.curso. Acesso em: 03 set. 2020.

BRASIL. Conselho Nacional De Saúde (CNS). Resolução nº 466, de 12 de dezembro de 2012. **Diário Oficial da União**, Brasília, DF, 13 jun. 2013. Seção 1, p. 59. Disponível em: <https://pesquisa.in.gov.br/imprensa/jsp/visualiza/index.jsp?data=13/06/2013&jornal=1&pagina=59&totalArquivos=140>. Acesso em: 06 jun. 2022.

BRASIL. Conselho Nacional De Saúde (CNS). Resolução Nº 510, de 07 de abril de 2016. **Diário Oficial da União**, Brasília, DF, 24 maio 2016. Seção 1, p. 44. Disponível em: <https://pesquisa.in.gov.br/imprensa/jsp/visualiza/index.jsp?jornal=1&data=24/05/2016&pagina=44>. Acesso em: 06 jun. 2022.

BRICKMAN, R. A visão do centro: políticas, desempenhos e paradoxos. In: SCHWARTZMAN, S.; CASTRO, C. M. (Orgs.). **Pesquisa universitária em questão**. São Paulo, SP: Ícone, 1986. p. 36-49.

BUDAPEST OPEN ACCESS INITIATIVE. **Budapest Open Access Initiative: Frequently Asked Questions**. Budapeste, 2002. Disponível em: <http://legacy.earlham.edu/~peters/fos/boaifaq.htm>. Acesso em: 27 jan. 2021.

BUDAPEST OPEN ACCESS INITIATIVE. **Read the Budapest open access initiative**. Budapeste, 2002. Disponível em: <https://www.budapestopenaccessinitiative.org/read>. Acesso em: 27 jan. 2021.

CAMPELLO, Bernadete. **Introdução ao controle bibliográfico**. 2. ed. Brasília: Briquet de Lemos/Livros, 2006. 94 p.

CARVALHO, E. R. S. **Diferenças na produção, compartilhamento e (re)uso de dados: a percepção de pesquisadores da química, antropologia e educação**. 2018. 173 f., il. Dissertação (Mestrado em Ciência da Informação) – Universidade de Brasília, Brasília, 2018.

CARVALHO, E. R. S.; LEITE, F. C. L. Diferenças na produção, no compartilhamento e no (re)uso de dados de pesquisa: a percepção de pesquisadores de Química, Antropologia e Educação. **Em Questão**, Porto Alegre, v. 25, n. 3, p. 321-347, set./dez. 2019. Disponível em: <https://www.seer.ufrgs.br/EmQuestao/article/view/89449/53591>. Acesso em: 03 mar. 2021.

CHARMAZ, K. **A construção da teoria fundamentada**: guia prático para análise qualitativa. Porto Alegre: Artmed, 2009. 272 p.

CHAWINGA, W. D.; ZINN, S. Research data management at a public university in Malawi: the role of “three hands”. **Library Management**, v. 41, n. 6/7, p. 467-485, 2020.

CHEN, X.; WU, M. Survey on the needs for chemistry research data management and sharing. **The Journal of Academic Librarianship**, v. 43, p. 346-353, 2017.

CHIWARE, E. R. T. Data librarianship in South African academic and research libraries: a survey. **Library Management**, v. 41, n. 6/7, p. 401-416, 2020.

COLLIS, J.; HUSSEY, R. **Business research**: A practical guide for undergraduate and postgraduate students. . ed. Houndmills, Basingstoke, Hampshire: New York: Palgrave MacMillan Higher Education, 2013. 376 p.

CONSORTIUM OF EUROPEAN SOCIAL SCIENCE DATA ARCHIVES (CESSDA). **Data in the social sciences**. Disponível em: <https://www.cessda.eu/Training/Training-Resourcques/Library/Data-Management-Expert-Guide/1.-Plan/Data-in-the-social-sciences>. Acesso em 26 maio 2021.

COSTA, M. P. **Fatores que influenciam a comunicação de dados de pesquisa sobre o vírus da zika, na perspectiva de pesquisadores**. Tese (Doutorado em Ciência da Informação) – Universidade de Brasília, Brasília, 2017. Disponível em: <https://repositorio.unb.br/handle/10482/23000>. Acesso em: 18 jun. 2021.

COSTA, S. M. S. Filosofia aberta, modelos de negócios e agências de fomento: elementos essenciais a uma discussão sobre o acesso aberto à informação científica. **Ciência da informação**, Brasília, v. 35, n. 2, p. 39-50, maio/agosto 2006. Disponível em: <https://www.scielo.br/pdf/ci/v35n2/a05v35n2.pdf>. Acesso em: 27 jan. 2021.

COSTA, S. M. S. Abordagens, estratégias e ferramentas para o acesso aberto via periódicos e repositórios institucionais em instituições acadêmicas brasileiras. **Liinc em Revista**, Rio de Janeiro, v. 4, n. 2, p. 218-232, set. 2008. Disponível em: <http://revista.ibict.br/liinc/article/view/3175/2840>. Acesso em: 03 set. 2020.

COX, A. M.; PINFIELD, S.; SMITH, J. Moving a brick building: UK libraries coping with research data management as a 'wicked' problem. **Journal of Librarianship and Information Science**, v. 48, n. 1, p. 3-17, 2016. Disponível em: <https://journals.sagepub.com/doi/pdf/10.1177/0961000614533717>. Acesso em: 17 jun. 2021.

CREATIVE COMMONS. **Sobre as licenças**. Disponível em: <https://creativecommons.org/licenses/?lang=pt>. Acesso em: 16 jun. 2021.

CRESWELL, J. W. **Investigação qualitativa e projeto de pesquisa**: escolhendo entre cinco abordagens. 3. ed. Porto Alegre: Penso, 2014.

CRESWELL, J. W. **Projeto de pesquisa**: métodos qualitativo, quantitativo e misto. 3. ed. Porto Alegre, RS: Artmed, 2010. 296 p. (Biblioteca Artmed. Métodos de Pesquisa).

CRIBB, J.; HARTOMO, T. S. **Open science**: sharing knowledge in the global century. Collingwood, Vic: CSIRO Publishing, 2010. 232 p.

DATA OBSERVATION NETWORK FOR EARTH (DATAONE). **Data management guide for public participation in scientific research**. 2013. 15 p. Disponível: <https://www.dataone.org/sites/all/documents/DataONE-PPSR-DataManagementGuide.pdf>. Acesso em: 03 set. 2020.

DE NEGRI, F.; CAVALCANTE, L. R.; ALVES, P. R. **Relações universidade-empresa no Brasil**: o papel da infraestrutura pública de pesquisa. Brasília; Rio de Janeiro: Ipea, 2013. 37 p.

DECLARAÇÃO DE BERLIM. **Declaração de Berlim sobre Acesso Livre ao Conhecimento nas Ciências e Humanidades**. Berlim. 2003. Disponível em: <https://repositorium.sdum.uminho.pt/about/DeclaracaoBerlim.htm>. Acesso em: 22 jan. 2021.

DIGITAL CURATION CENTRE (DCC). **ABCD - Access to Biological Collection Data**. Disponível em: <https://www.dcc.ac.uk/resources/metadata-standards/abcd-access-biological-collection-data>. Acesso em: 17 jun. 2021.

DIGITAL CURATION CENTRE (DCC). **AgMES - Agricultural Metadata Element Set**. Disponível em: <https://www.dcc.ac.uk/resources/metadata-standards/agmes-agricultural-metadata-element-set>. Acesso em: 17 jun. 2021.

DIGITAL CURATION CENTRE (DCC). **AVM - Astronomy Visualization Metadata**. Disponível em: <https://www.dcc.ac.uk/resources/metadata-standards/avm-astronomy-visualization-metadata>. Acesso em: 17 jun. 2021.

DIGITAL CURATION CENTRE (DCC). **CERIF - Common European Research Information Format**. Disponível em: <https://www.dcc.ac.uk/resources/metadata-standards/cerif-common-european-research-information-format>. Acesso em: 17 jun. 2021.

DIGITAL CURATION CENTRE (DCC). **CF (Climate and Forecast) Metadata Conventions**. Disponível em: <https://www.dcc.ac.uk/resources/metadata-standards/cf-climate-and-forecast-metadata-conventions>. Acesso em: 17 jun. 2021.

DIGITAL CURATION CENTRE (DCC). **CIF - Crystallographic Information Framework**. Disponível em: <https://www.dcc.ac.uk/resources/metadata-standards/cif-crystallographic-information-framework>. Acesso em: 17 jun. 2021.

DIGITAL CURATION CENTRE (DCC). **Darwin Core**. Disponível em: <https://www.dcc.ac.uk/resources/metadata-standards/darwin-core>. Acesso em: 17 jun. 2021.

DIGITAL CURATION CENTRE (DCC). **Data Package**. Disponível em: <https://www.dcc.ac.uk/resources/metadata-standards/data-package>. Acesso em: 17 jun. 2021.

DIGITAL CURATION CENTRE (DCC). **DataCite Metadata Schema**. Disponível em: <https://www.dcc.ac.uk/resources/metadata-standards/datacite-metadata-schema>. Acesso em: 17 jun. 2021.

DIGITAL CURATION CENTRE (DCC). **DDI - Data Documentation Initiative**. Disponível em: <https://www.dcc.ac.uk/resources/metadata-standards/ddi-data-documentation-initiative>. Acesso em: 17 jun. 2021.

DIGITAL CURATION CENTRE (DCC). **DIF - Directory Interchange Format**. Disponível em: <https://www.dcc.ac.uk/resources/metadata-standards/dif-directory-interchange-format>. Acesso em: 17 jun. 2021.

DIGITAL CURATION CENTRE (DCC). **Disciplinary Metadata**. Disponível em: <https://www.dcc.ac.uk/guidance/standards/metadata>. Acesso em: 17 jun. 2021.

DIGITAL CURATION CENTRE (DCC). **Dublin Core**. Disponível em: <https://www.dcc.ac.uk/resources/metadata-standards/dublin-core>. Acesso em: 17 jun. 2021.

DIGITAL CURATION CENTRE (DCC). **EML - Ecological Metadata Language**. Disponível em: <https://www.dcc.ac.uk/resources/metadata-standards/eml-ecological-metadata-language>. Acesso em: 17 jun. 2021.

DIGITAL CURATION CENTRE (DCC). **Genome Metadata**. Disponível em: <https://www.dcc.ac.uk/resources/metadata-standards/genome-metadata>. Acesso em: 17 jun. 2021.

DIGITAL CURATION CENTRE (DCC). **International Virtual Observatory Alliance Technical Specifications (IVOA)**. Disponível em: <https://www.dcc.ac.uk/resources/metadata-standards/international-virtual-observatory-alliance-technical-specifications>. Acesso em: 17 jun. 2021.

DIGITAL CURATION CENTRE (DCC). **ISO 19115**. Disponível em: <https://www.dcc.ac.uk/resources/metadata-standards/iso-19115>. Acesso em: 17 jun. 2021.

DIGITAL CURATION CENTRE (DCC). **MIDAS-Heritage**. Disponível em: <https://www.dcc.ac.uk/resources/metadata-standards/midas-heritage>. Acesso em: 17 jun. 2021.

DIGITAL CURATION CENTRE (DCC). **NeXus**. Disponível em: <https://www.dcc.ac.uk/resources/metadata-standards/nexus>. Acesso em: 17 jun. 2021.

DIGITAL CURATION CENTRE (DCC). **QuDEx - Qualitative Data Exchange Format**. Disponível em: <https://www.dcc.ac.uk/resources/metadata-standards/qudex-qualitative-data-exchange-format>. Acesso em: 17 jun. 2021.

DIGITAL CURATION CENTRE (DCC). **SDMX - Statistical Data and Metadata Exchange**. Disponível em: <https://www.dcc.ac.uk/resources/metadata-standards/sdmx-statistical-data-and-metadata-exchange>. Acesso em: 17 jun. 2021.

DIGITAL PANTHEON. **Overview**. Disponível em: <http://repository.edition-topoi.org/collection/BDPP/overview>. Acesso em: 17 jun. 2021.

DUARTE, J. Entrevista em profundidade. In: DUARTE, Jorge; BARROS, Antonio (Orgs.). **Métodos e técnicas de pesquisa em comunicação**. 2. ed. São Paulo: Atlas, 2011. cap. 4. p. 62-83.

DUNIE, M. The importance of research data management: the value of electronic laboratory notebooks in the management of data integrity and data availability. **Information Services & Use**, v. 37, p. 355-359, 2017. Disponível em: <https://content.iospress.com/download/information-services-and-use/isu843?id=information-services-and-use%2Fisu843>. Acesso em: 03 set. 2020.

EUROPEAN COMMISSION. Directorate-General for Research and Innovation. **Study on open science: impact, implications and policy options**. Luxembourg: Publications Office of the European Union, 2015. Disponível em: https://ec.europa.eu/research/innovation-union/pdf/expert-groups/rise/study_on_open_science-impact_implications_and_policy_options-salmi_072015.pdf. Acesso em 03 set. 2020.

FACILITATE OPEN SCIENCE TRAINING FOR EUROPEAN RESEARCH (FOSTER PLUS).
About FOSTER. Disponível em:
<https://www.fosteropenscience.eu/about#theprevproject>. Acesso em: 03 set. 2020.

FACILITATE OPEN SCIENCE TRAINING FOR EUROPEAN RESEARCH (FOSTER PLUS).
Open Science Taxonomy. Disponível em: <https://www.fosteropenscience.eu/foster-taxonomy/open-science>. Acesso em: 03 set. 2020.

FECHER, B; FRIESIKE, S. Open Science: one term, five schools of thought. In: BARTLING, S.; FRIESIKE, S (Ed.). **Opening Science: the evolving guide on how the Internet is changing research, collaboration and scholarly publishing.** London: Springer Open, 2014, p. 17-47. Disponível em:
<https://link.springer.com/content/pdf/10.1007%2F978-3-319-00026-8.pdf>. Acesso em: 03 set. 2020.

FEDERER, L. Research data management in the age of big data: roles and opportunities for librarians. **Information Services & Use**, v. 36, n. 1-2, p. 35-43, 2016. Disponível em: <https://content.iospress.com/articles/information-services-and-use/isu797>. Acesso em: 03 set. 2020.

FEDERER, L. Defining data librarianship: a survey of competencies, skills, and training. **Journal of the Medical Library Association: JMLA**, v. 106, n. 3, p. 294-303, jul. 2018. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6013124/pdf/jmla-106-294.pdf>. Acesso em: 16 fev. 2021.

FERREIRA, A. B. de H. **Novo Aurélio século xxi: o dicionário da língua portuguesa.** 3. ed. Rio de Janeiro: Nova Fronteira, 2000. 2128 p.

FIGSHARE. **What is Figshare?** Disponível em: <https://help.figshare.com/article/what-is-figshare>. Acesso em: 17 jun. 2021.

FLICK, U. **Desenho da pesquisa qualitativa.** Porto Alegre: Bookman, 2009. 164 p.

FUNDAÇÃO DE AMPARO À PESQUISA DO ESTADO DE SÃO PAULO (FAPESP). **Covid-19 DataSharing/BR.** Disponível em:
<https://repositoriodatasharingfapesp.uspdigital.usp.br/>. Acesso em: 16 jun. 2021.

GALVÃO, M. C. B.; RICARTE, I. L. M. Revisão sistemática da literatura: conceituação, produção e publicação. **Logeion: Filosofia da informação**, v. 6, n. 1, p. 57-73, 2019. Disponível em: <http://revista.ibict.br/fiin/article/view/4835>. Acesso em: 17 jun. 2021.

GASKELL, G. Entrevistas individuais e grupais. In: BAUER, Martin W.; GASKELL, George (Org.). **Pesquisa qualitativa com texto, imagem e som: um manual prático.** 4. ed. Petrópolis: Vozes, 2005. cap. 3. p. 64-89.

GERMAN FEDERATION FOR BIOLOGICAL DATA (GFBIO). **Data life cycle**. Disponível em: <https://www.gfbio.org/pt/training/materials/data-lifecycle>. Acesso em: 03 set. 2020.

GERMAN FEDERATION FOR BIOLOGICAL DATA (GFBIO). **What is the German Federation for Biological Data (GFBio)?** Disponível em: <https://www.gfbio.org/pt/support/faqs>. Acesso em: 03 set. 2020.

GLOBAL NATURAL PRODUCTS SOCIAL MOLECULAR NETWORKING (GNPS). **GNPS**. Disponível em: <https://gnps.ucsd.edu/ProteoSAFe/static/gnps-splash.jsp>. Acesso em: 18 nov. 2021.

FERREIRA, A. A; GUIMARÃES, E. R.; CONTADOR, J. C. Patente como instrumento competitivo e como fonte de informação tecnológica. **Gest. Prod.**, São Carlos, v. 16, n. 2, p. 209-221, abr./jun. 2009. Disponível em: <https://www.scielo.br/j/gp/a/HjrgQJZqTkZvVhZp3fJT8hw/?format=pdf&lang=pt>. Acesso em: 25 out. 2022.

GILMORE; R. O. *et al.* Advances in Engineering Education: transforming education research through open video data sharing. **Advances In Engineering Education**, v. 5, n. 2, p. 1-17, spring 2016. Disponível em: <https://advances.asee.org/wp-content/uploads/vol05/issue02/Papers/AEE-18-Gilmore.pdf>. Acesso em: 24 out. 2022.

GUIMARÃES, A. J. R.; BEZERRA, C. A. Gestão de dados: uma abordagem bibliométrica. **Perspectivas em Ciência da Informação**, v. 24, n. 4, p. 171-186, out./dez. 2019. Disponível em: <https://www.scielo.br/j/pci/a/3gGq6KQNKp7Mqx7sWsfkj6N/?format=pdf&lang=pt>. Acesso em: 18 jun. 2021.

HEAVY.AI. **Geodata definition**. Disponível em: <https://www.heavy.ai/technical-glossary/geodata>. Acesso em: 11 nov. 2022.

HERNÁNDEZ SAMPIERI, R.; FERNÁNDEZ COLLADO, C.; BAPTISTA LUCIO, P. **Metodologia de pesquisa**. 3. ed. São Paulo: McGraw-Hill, 2006. 583 p.

HIGMAN, R.; PINFIELD, S. Research data management and openness: the role of data sharing in developing institutional policies and practices. **Program**, v. 49, n. 4, p. 364-381, 2015. Disponível em: [//eprints.whiterose.ac.uk/89888/1/RDM%20and%20Openness%20revised%20Program%201-3%20final%20submitted%20version%20inc%20figs%20and%20tabs%20WRRO.pdf](https://eprints.whiterose.ac.uk/89888/1/RDM%20and%20Openness%20revised%20Program%201-3%20final%20submitted%20version%20inc%20figs%20and%20tabs%20WRRO.pdf). Acesso em: 17 jun. 2021.

HIGMAN, R; BANGERT, D.; JONES, S. Three camps, one destination: the intersections of research data management, FAIR and Open. **Insights**, v. 32, n. 1, p. 1-9, 2019. Disponível em: <https://insights.uksg.org/articles/10.1629/uksg.468/#>. Acesso em: 17 jun. 2021.

HOUAISS, A.; VILLAR, M. S.; FRANCO, F. M. M. **Dicionário Houaiss da língua portuguesa**. Rio de Janeiro: Objetiva, 2009. 1986 p.

INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA (IBICT). **Rede Brasileira de Serviços de Preservação Digital – CARINIANA**. Disponível em: <https://ibict.br/tecnologias-para-informacao/cariniana#apresentacao>. Acesso em: 17 jun. 2021.

INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA (IBICT). **Manifesto de Acesso Aberto a Dados da Pesquisa Brasileira para Ciência Cidadã**. 2017. Disponível em: <https://ibict.br/sala-de-imprensa/noticias/item/478-ibict-lanca-manifesto-de-acesso-aberto-a-dados-da-pesquisa-brasileira-para-ciencia-cidada>. Acesso em 16 jan. 2021.

INTER-UNIVERSITY CONSORTIUM FOR POLITICAL AND SOCIAL RESEARCH (ICPSR). **Guide to social Science data preparation and archiving**. Michigan: Institute for Social Research University. 2012. 45 p. Disponível em: <https://deepblue.lib.umich.edu/bitstream/handle/2027.42/134032/dataprep.pdf?sequence=1&isAllowed=y>. Acesso em: 03 set. 2020.

JAPIASSU, H. **Interdisciplinaridade e patologia do saber**. Rio de Janeiro: Imago, 1976. 220 P.

JISC. **Research data lifecycle**. Disponível em: <https://rdmtoolkit.jisc.ac.uk/research-data-lifecycle/>. Acesso em 03 set. 2020.

JOMIER, J. Open science – towards reproducible research. **Information Services & Use**, v. 37, n. 3, p. 361–367, 2017. Disponível em: <https://content.iospress.com/download/information-services-and-use/isu846?id=information-services-and-use%2Fisu846>. Acesso em: 03 set. 2020.

KNORR-CETINA, Karin. A comunicação na ciência. In: GIL, Fernando (Ed.). **A Ciência tal qual se Faz**. Lisbon: Edições João Sá da Costa, 1999. p. 375-393.

LEITE, F. C. L. **Como gerenciar e ampliar a visibilidade da informação científica brasileira: repositórios institucionais de acesso aberto**. Brasília: Ibict, 2009. 120 p. Disponível em: <https://livroaberto.ibict.br/bitstream/1/775/4/Como%20gerenciar%20e%20ampliar%20a%20visibilidade%20da%20informa%20c3%a7%20a3o%20cient%20c3%adfica%20brasileira.pdf>. Acesso em: 26 jan. 2021.

LEITE, F. C. L. Práticas de busca, acesso e disseminação da informação científica de pesquisadores de diferentes áreas do conhecimento. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 13., 2012, Rio de Janeiro. **Anais...** Rio de Janeiro: ANCIB, 2012. Disponível em:

<http://enancib.ibict.br/index.php/enancib/xiiienancib/paper/viewFile/3825/2948>. Acesso em: 03 set. 2020.

LÉXICO: DICIONÁRIO DE PORTUGUÊS ONLINE. **Significado de diretrizes**. Disponível em: <https://www.lexico.pt/diretrizes/>. Acesso em: 11 set. 2020.

MANIFIESTO BIBLIOTECARIO POR LA CIENCIA ABIERTA EN AMÉRICA LATINA. Bogotá, 2019. Versión en Español. Disponível em: <https://docutopia.tupale.co/bibliotecariosalsenado-manifiestobibliotecario2019>. Acesso em: 27 jan. 2021.

MARTINEZ-URIBE, L.; FERNÁNDEZ, P. Servicios de datos: función estratégica de las bibliotecas del siglo XXI. **El profesional de la información**, v. 24, n. 2, p. 193-199, 2015. Disponível em: <http://eprints.rclis.org/39578/1/13.pdf>. Acesso em: 16 fev. 2021.

MEADOWS, A. J. **A comunicação científica**. Brasília: Briquet De Lemos, 1999. 268 p.

MELERO, Remedios. Acceso abierto a las publicaciones científicas: definición, recursos, copyright e impacto. **El profesional de la información**, v. 14, n. 4, p. 255-266, jul.-ago. 2005. Disponível em: <http://eprints.rclis.org/6571/>. Acesso: 27 jan. 2021.

MEY, Eliane Serrão Alves; SILVEIRA, Naira Christofolletti. **Catálogo no plural**. Brasília: Briquet de Lemos/Livros, 2009. ix, 217 p.

MOLLOY, J. C. The Open Knowledge Foundation: Open Data Means Better Science. **PLoS Biol**, v. 9, n. 12, p. 1-4, dec. 2011.

MUELLER, S. P. M. A comunicação científica e o movimento de acesso livre ao conhecimento. **Ciência da informação**, Brasília, v. 35, n. 2, p. 27-38, maio/ago. 2006. Disponível em: <https://www.scielo.br/pdf/ci/v35n2/a04v35n2.pdf>. Acesso em: 27 jan. 2021.

MUELLER, S. P. M. A publicação da ciência: áreas científicas e seus canais preferenciais. **DataGramZero: Revista de Ciência da Informação**, v. 6, n. 1, fev. 2005. Disponível em: <http://repositorio.unb.br/handle/10482/980>. Acesso em 03 set. 2020.

MURRAY-RUST, P. et al. **Panton principles: Open Data in Science**. 2010. Disponível em: <https://pantonprinciples.org/>. Acesso em 16 jan. 2021.

NASCIMENTO, L. P. **Elaboração de projetos de pesquisa**: monografia, dissertação, tese e estudo de caso, com base em metodologia científica. São Paulo: Cengage Learning, 2012.

NATIONAL INSTITUTES OF HEALTH (NIH). **Data Repository**. Disponível em: <https://nnlm.gov/data/thesaurus/data-repository>. Acesso em: 17 jun. 2021.

NATIONAL LIBRARY OF MEDICINE. **GenBank Overview**. Disponível em: <https://www.ncbi.nlm.nih.gov/genbank/>. Acesso em: 18 nov. 2022.

NOONAN, D.; CHUTE, T. Data Curation and the University Archives. **The American Archivist**, v. 77, n. 1, p. 201-240, apr. 2014.

OBSERVATÓRIO DO CONHECIMENTO. **Balço anual do orçamento do conhecimento: 2021**. Disponível em: https://observatoriodoconhecimento.org.br/wp-content/uploads/2022/05/balanco-anual-orcamento-do-conhecimento-2021_compactado.pdf. Acesso em: 25 out. 2022.

OPEN KNOWLEDGE FOUNDATION (OKF). **Open Data Handbook: O que são dados abertos?** Disponível em: http://opendatahandbook.org/guide/pt_BR/what-is-open-data/. Acesso em: 03 set. 2020.

OPEN KNOWLEDGE FOUNDATION (OKF). **Open Data Handbook: Why open data?** Disponível em: <https://opendatahandbook.org/guide/en/why-open-data/>. Acesso em: 03 set. 2020.

OPEN KNOWLEDGE FOUNDATION (OKF). **What kinds of open data?** Disponível em: <https://okfn.org/opendata/>. Acesso em: 03 set. 2020.

OPEN KNOWLEDGE FOUNDATION. **Open Data Commons Attribution License (ODC-By) v1.0**. Disponível em: <https://opendatacommons.org/licenses/by/1-0/>. Acesso em: 16 jun. 2021.

OPEN KNOWLEDGE FOUNDATION. **Open Data Commons Open Database License (ODbL) v1.0**. Disponível em: <https://opendatacommons.org/licenses/odbl/1-0/>. Acesso em: 16 jun. 2021.

OPEN KNOWLEDGE FOUNDATION. **Open Data Commons Public Domain Dedication and License (PDDL) v1.0**. Disponível em: <https://opendatacommons.org/licenses/pddl/1-0/>. Acesso em: 16 jun. 2021.

OPEN KNOWLEDGE FOUNDATION. **Open Data Commons: legal tools for open data**. Disponível em: <https://opendatacommons.org/>. Acesso em: 16 jun. 2021.

OPEN RESPONSIBLE RESEARCH AND INNOVATION TO FURTHER OUTSTANDING KNOWLEDGE (ORIU). **About ORION**. Disponível em: <https://www.orion-openscience.eu/about>. Acesso em: 03 set. 2020.

OPEN RESPONSIBLE RESEARCH AND INNOVATION TO FURTHER OUTSTANDING KNOWLEDGE (ORIIUM). **What is Open Science?** Disponível em: <https://www.orion-openscience.eu/resources/open-science>. Acesso em: 03 set. 2020.

OPENAIRE. **What-is-metadata**. Disponível em: <https://www.openaire.eu/what-is-metadata>. Acesso em: 17 jun. 2021.

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT (OECD). **Principles and guidelines for access to research data from public funding**. France: OECD Publications, 2007. Disponível em: <https://www.oecd.org/sti/sci-tech/38500813.pdf>. Acesso em: 03 set. 2020.

ORGANIZAÇÃO DAS NAÇÕES UNIDAS PARA A EDUCAÇÃO, A CIÊNCIA E A CULTURA (UNESCO). **Guidelines for a Research Data Management Plan**. Paris: IOC, 2016. 14 p. Disponível em: <https://unesdoc.unesco.org/ark:/48223/pf0000256544>. Acesso em 03 jun. 2020.

PACHECO, J. T. S.; ABREU, N. O.; BERNARDINO, C. G. Culturas disciplinares: influências na produção do gênero artigo acadêmico. **Revista linguagem em foco**, Ceará, v. 10, n. 1, p. 71-82, 2018. Disponível em: <https://revistas.uece.br/index.php/linguagememfoco/article/view/1190/1006>. Acesso em: 16 jun. 2021.

PAMPEL, H. *et al.* Making Research Data Repositories Visible: The re3data.org Registry. **PLoS ONE**, v. 8, n. 11, p. 1-10, nov. 2013. Disponível em: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0078080>. Acesso em: 17 jun. 2021.

PARÉ G; KITSIOU S. Methods for literature reviews. In: LAU F; KUZIEWSKY C. (Ed.). **Handbook of ehealth evaluation: an evidence-based approach**. Victoria, BC: University of Victoria, 2016. cap. 9. p. 157-179. Disponível em: https://www.ncbi.nlm.nih.gov/books/NBK481590/pdf/Bookshelf_NBK481590.pdf. Acesso em: 03 set. 2020.

PATEL, D. Research data management: a conceptual framework. **Library Review**, v. 65, n. 4-5, p. 226-241, 2016.

PERRIER, L; BARNES, L. Developing research data management services and support for researchers: a mixed methods study. **Partnership: The Canadian Journal of Library and Information Practice and Research**, v. 13, n. 1, p. 1-23, 2018. Disponível em: <https://www.semanticscholar.org/paper/Developing-Research-Data-Management-Services-and-A-Perrier-Barnes/3b1a9baaa2202f3ca6d415724849eb239a92b689>. Acesso em: 17 jun. 2021.

PINFIELD, S. Journals and repositories: an evolving relationship?. **Learned publishing**, v. 22, n. 3, p. 165-175, 2009. Disponível em:

<https://onlinelibrary.wiley.com/doi/pdfdirect/10.1087/2009302>. Acesso em: 17 jun. 2021.

PIRACHA, H. A; AMEEN, K. Research Data Management Practices of Faculty Members. **Pakistan Journal of Information Management and Libraries**, v. 20, p. 60-75, 2018. Disponível em:

<http://journals.pu.edu.pk/journals/index.php/pjimpl/article/viewFile/1321/667>. Acesso em: 17 jun. 2021.

PLOMP, E. *et al.* Cultural obstacles to research data management and sharing at TU Delft. **Insights**, v. 32, n. 1, p. 1-11, 2019. Disponível em:

<https://insights.uksg.org/article/10.1629/uksg.484/>. Acesso em: 17 jun. 2021.

POMBO, O. Da classificação dos seres à classificação dos saberes. **Revista da Biblioteca Nacional de Lisboa**, Lisboa, v. 2, p. 19-33, 1998. Disponível em:

https://www.researchgate.net/profile/Olga_Pombo/publication/255612117_Da_classificacao_dos_seres_a_classificacao_dos_saberes/links/0a85e52dcbb59e738b000000/Da-classificacao-dos-seres-a-classificacao-dos-saberes.pdf. Acesso em: 03 set. 2020.

POUPART, J. A entrevista de tipo qualitativo: considerações epistemológicas, teóricas e metodológicas. In: _____ *et al.* **A pesquisa qualitativa: enfoques epistemológicos e metodológicos**. Petrópolis, RJ: Vozes, 2008. 464 p. (Coleção Sociologia)

PROQUEST. **LISA: Library and Information Science Abstracts**. Disponível em:

<http://proquest.libguides.com/lisa>. Acesso em: 03 set. 2020.

RE3DATA.ORG. **FAQ - What is re3data?** Disponível em: <https://www.re3data.org/faq>. Acesso em: 17 jun. 2021.

RESEARCH COUNCILS UK. **Concordat on Open Research Data**. 2016. 24 p. Disponível em: <https://www.ukri.org/wp-content/uploads/2020/10/UKRI-020920-ConcordatonOpenResearchData.pdf>. Acesso em: 24 out. 2022.

RICE, R. Supporting research data management and open science in academic libraries: a data librarian's view. **Mitteilungen der Vereinigung Österreichischer Bibliothekarinnen und Bibliothekare**, v. 72, n. 2, p. 263-273, 2019.

RODRIGUES, E. *et al.* **Os repositórios de dados científicos: estado da arte**. Portugal: Universidade do Porto, 2010. 54 p. Disponível em: <https://repositorio-aberto.up.pt/bitstream/10216/23806/2/44632.pdf>. Acesso em: 17 jun. 2021.

RÜEGG *et al.* Completing the data life cycle: using information management in macrosystems ecology research. **Frontiers in Ecology Environment**, v. 12, n. 1, p. 24-30, 2014. Disponível em:

<https://esajournals.onlinelibrary.wiley.com/doi/pdf/10.1890/120375>. Acesso em: 03 set. 2020.

SAEED; S.; ALI, P. M. N. Research data management and data sharing among research scholars of Life Sciences and Social Sciences. **DESIDOC Journal of Library & Information Technology**, Índia, v. 39, n. 6, p. 290-299, 2019.

SALES, L. F. *et al.* Competências dos bibliotecários na gestão dos dados de pesquisa. Competências dos bibliotecários na gestão dos dados de pesquisa. **Ciência da Informação**, Brasília, v. 48, n. 3 (Supl.), p. 303-313, 2019. Disponível em: <http://revista.ibict.br/ciinf/article/view/4973/4458>. Acesso em: 16 fev. 2021.

SALES; L. F.; SAYÃO, L. F. A grande e a pequena ciência: análise das diferenças na gestão de dados de pesquisa. **Informação & Sociedade: Estudos**, João Pessoa, v. 29, n. 3, p. 151-170, jul./set. 2019. Disponível em: <https://periodicos.ufpb.br/ojs/index.php/ies/article/view/47615/27371>. Acesso em: 03 mar. 2021.

SANCHEZ, F. A.; SILVA, N. B. P.; VECHIATO, F. L. Padrões de metadados para representação e organização da informação em repositórios de dados de pesquisa. **Informação & Tecnologia (ITEC)**, Marília/João Pessoa, v. 5, n. 1, p. 51-37, jan./jun. 2018.

SAYÃO, L. F. Uma outra face dos metadados: informações para a gestão da preservação digital. **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação**, Florianópolis, v. 15, n. 30, p. 1-31, 2010. Disponível em: <https://www.redalyc.org/pdf/147/14716919002.pdf>. Acesso em: 17 jun. 2021.

SAYÃO, L. F.; SALES, L. F. Algumas considerações sobre os repositórios digitais de dados de pesquisa. **Informação & Informação**, Londrina, v. 21, n. 2, p. 90-115, 2016. Disponível em: <http://www.uel.br/revistas/uel/index.php/informacao/article/view/27939>. Acesso em: 17 jun. 2021.

SAYÃO, L. F.; SALES, L. F. Curadoria digital: um novo patamar para preservação de dados digitais de pesquisa. **Inf. & Soc.:Est.**, João Pessoa, v. 22, n. 3, p. 179-191, set./dez. 2012. Disponível em: <https://www.marilia.unesp.br/Home/Instituicao/Docentes/EdbertoFerneda/curadoria-digital---sayao.pdf>. Acesso em: 02 jan. 2023.

SAYÃO, L. F.; SALES, L. F. Dados abertos de pesquisa: ampliando o conceito de acesso livre. **Revista Eletrônica de Comunicação, Informação & Inovação em Saúde**, Rio de Janeiro, v. 8, n. 2, p. 76-92, jun. 2014. Disponível em: <https://www.reciis.icict.fiocruz.br/index.php/reciis/article/view/611/1252>. Acesso em: 03 set. 2020.

SAYÃO, L. F.; SALES, L. F. **Guia de gestão de dados de pesquisa para bibliotecários e pesquisadores**. Rio de Janeiro: CNEN/ IEN, 2015. 90 p. Disponível em: http://www.cnen.gov.br/images/CIN/PDFs/GUIA_DE_DADOS_DE_PESQUISA.pdf. Acesso

em: 03 set. 2020.

SCHWARTZMAN, S.; CASTRO, C. M. (Orgs.). **Pesquisa universitária em questão**. São Paulo, SP: Ícone, 1986. 232 p.

SCIELO. **SciELO**: Modelo de publicação eletrônica para países em desenvolvimento. SciELO, 2019. Disponível em: https://wp.scielo.org/wp-content/uploads/Modelo_SciELO.pdf. Acesso em: 30 dez. 2022.

SEKARAN, U. **Research methods for business: a skill-building approach**. 4. ed. New York: Wiley, c2003.

SEMELER, A. R.; PINTO, A. L. Data librarianship as a field study. **Transinformação**, v. 32, 2020. Disponível em: <https://www.scielo.br/pdf/tinf/v32/2318-0889-tinf-32-e200034.pdf>. Acesso em: 16 fev. 2021.

SEMELER, A. R.; PINTO, A. L. Os diferentes conceitos de dados de pesquisa na abordagem da biblioteconomia de dados. **Ciência da Informação**, Brasília, v. 48, n. 1, p. 113-129, jan./abr. 2019. Disponível em: <http://revista.ibict.br/ciinf/article/view/4461/4102>. Acesso em: 16 fev. 2021.

SEMELER, A. R.; PINTO, A. L.; ROZADOS, H. B. F. Data science in data librarianship: core competencies of a data librarian. **Journal of Librarianship and Information Science**, v. 51, n. 3, p. 771-780, 2019.

SEVERINO, A. J. Educação e universidade: conhecimento e construção da cidadania. **Interface: Comunicação, Saúde, Educação**, São Paulo, v. 6, n. 10, p. 117-124, fev. 2002. Disponível em: <https://www.scielo.br/j/icse/a/dcRwJ5SSQW6Z43pGWpKyMkc/?lang=pt>. Acesso em: 16 jun. 2021.

SIEMENS, L.; LIU, Y.; SMITH, J. Mapping disciplinary differences and equity of academic control to create a space for collaboration. **Canadian Journal of Higher Education**, v. 44, n. 2, p. 49-67, 2014. Disponível em: <https://journals.sfu.ca/cjhe/index.php/cjhe/article/view/183760/185211>. Acesso em: 08 nov. 2022.

SHIPMAN, J. P.; TANG, R. The collaborative creation of a Research Data Management Librarian Academy (RDMLA). **Information Services & Use**, v. 39, n. 3, p. 243-247, 2019. Disponível em: <https://content.iospress.com/articles/information-services-and-use/isu190050>. Acesso em: 17 jun. 2021.

SILVA, F. C. C. **Gestão de dados científicos**. Rio de Janeiro: Interciência, 2019. 128 p.

SILVA, F. C. C. O papel dos bibliotecários na gestão de dados científicos. RDBCI: **Revista Digital de Biblioteconomia e Ciência da Informação**, v. 14, n. 3, p. 387-406, set./dez. 2016. Disponível em: <http://docplayer.com.br/156569618-Papel-dos-bibliotecarios-na-gestao-de-dados-cientificos.html>. Acesso em: 25 mar. 2021.

SINDICATO NACIONAL DOS DOCENTES DAS INSTITUIÇÕES DE ENSINO SUPERIOR (Andes). **Pesquisa nacional luta para sobreviver asfixiada por cortes orçamentários**. Disponível em: <https://www.andes.org.br/conteudos/noticia/pesquisa-nacional-luta-para-sobreviver-asfixiada-por-cortes-orcamentarios1>. Acesso em: 20 nov. 2022.

SIQUEIRA, J. C. O conceito classificação: uma abordagem histórica e epistemológica. **Revista Brasileira de Biblioteconomia e Documentação**, São Paulo, v. 6, n. 1, p. 37-49, jan./jun. 2010. Disponível em: <https://rbbd.febab.org.br/rbbd/article/viewFile/137/167>. Acesso em: 03 set. 2020.

SOUZA, S. S. **CDU: como entender e utilizar a 2ª Edição-Padrão Internacional em Língua Portuguesa**. 2. ed. Brasília: Thesaurus, 2010. 163 p.

STRAUSS, A.; CORBIN, J. **Pesquisa qualitativa: técnicas e procedimentos para o desenvolvimento de teoria fundamentada**. 2. ed. Porto Alegre: Artmed, 2008.

SUBER, P. **Acceso Abierto**. Traducción de Remedios Melero. Toluca, Estado de México: Universidad Autónoma del Estado de México, 2015. 265 p. (Colección Cuadernos Institucionales). Disponível em: <https://digital.csic.es/bitstream/10261/121428/1/ACCESO%20ABIERTO.pdf>. Acesso em: 27 jan. 2021.

SURKIS, A.; READ, K. Research data management. **J Med Libr Assoc.**, v. 103, n. 3, p. 154-156, jul. 2015. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4511058/pdf/mlab-103-03-154.pdf>. Acesso em: 03 set. 2020.

TECHNICAL UNIVERSITY OF DENMARK DTU. **Guide to data management**. Disponível em: <https://www.bibliotek.dtu.dk/english/servicemenu/publish/research-data/guide>. Acesso em 07 set. 2020.

THOMAS, C. V. L.; URBAN, R. J. What do data librarians think of the MLIS? Professionals' perceptions of knowledge transfer, trends, and challenges. **College & Research Libraries**, v. 79, n. 3, p. 401-423, apr. 2018.

TOMANIK, E. A. **O olhar no espelho: "conversas" sobre a pesquisa em ciências sociais**. 2. ed. rev. Maringá, PR: Editora da Universidade Estadual de Maringá, 2004. 239 p.

UK DATA ARCHIVE. **About us**. Disponível em: <https://www.data-archive.ac.uk/about/>. Acesso em: 18 jun. 2021.

UK DATA ARCHIVE. **Research data lifecycle**. Disponível em: <https://www.youtube.com/watch?v=-wjFMMQD3UA>. Acesso em: 03 set. 2020.

UNITED STATES GEOLOGICAL SURVEY (USGS). **The United States Geological Survey Science Data Lifecycle Model**. U.S. Geological Survey, Reston, Virginia: 2013, 12 p. Disponível em: <https://pubs.usgs.gov/of/2013/1265/pdf/of2013-1265.pdf>. Acesso em: 03 set. 2020.

UNIVERSIDADE DE BRISTOL. **The data.bris Research Data Repository. About**. Disponível em: <https://data.bris.ac.uk/data/about>. Acesso em: 17 jun. 2021.

UNIVERSIDADE DE LEEDS. **What is research data?** Disponível em: https://library.leeds.ac.uk/info/14062/research_data_management/61/research_data_management_explained. Acesso em: 03 set. 2020.

UNIVERSIDADE DE MELBOURNE. **Management of Research Data and Records Policy**. Melbourne: Universidade de Melbourne, 2013. 6 p. Disponível em: <https://policy.unimelb.edu.au/MPF1242>. Acesso em: 03 set. 2020.

UNIVERSIDADE DE NEW ENGLAND. **Metadata for Research Data**. Disponível em: <https://www.une.edu.au/research/digital-research-support/research-data-management/metadata-for-research-data>. Acesso em: 17 jun. 2021.

UNIVERSIDADE DE NOTTINGHAM. **Research Data Management**. Disponível em: <https://www.nottingham.ac.uk/fabs/rgs/research-data-management/what-is-research-data.aspx>. Acesso em: 03 set. 2020.

UNIVERSIDADE DE OULU. **Research Data Guide - Documentation and metadata**. Disponível em: https://libguides oulu.fi/Researchdata/Data_documentation. Acesso em: 17 jun. 2021.

UNIVERSIDADE DE SÃO PAULO. **Dados de Pesquisa**. Disponível em: <http://www.sibi.usp.br/apoio-pesquisador/dados-pesquisa/>. Acesso em 07 set. 2020.

UNIVERSIDADE DE VIRGINIA. **Steps in the data life cycle**. Disponível em: <https://data.library.virginia.edu/data-management/lifecycle/>. Acesso em: 03 set. 2020.

UNIVERSIDADE DE YALE. **What is research data?** Disponível em: <https://guides.library.yale.edu/datamanagement>. Acesso em 03 set. 2020.

UNIVERSIDADE DO MINHO. **Sobre Repositórios OA**. Disponível em: https://openaccess.sdum.uminho.pt/?page_id=348. Acesso em: 17 jun. 2021.

UNIVERSIDADE FEDERAL DE PELOTAS. Pró-Reitoria de Pesquisa, Pós-Graduação e Inovação. Coordenação de Inovação Tecnológica. **Guia de orientação sobre:**

propriedade intelectual e patentes de invenção. 2018. 26 p. Disponível em: https://wp.ufpel.edu.br/epitte/files/2018/11/Guia-de-Propriedade-Intelectual_Ver.-4.0_05_11_2018.pdf?file=2018/11/Guia-de-Propriedade-Intelectual_Ver.-4.0_05_11_2018.pdf. Acesso em: 20 nov. 2022.

UNIVERSITY COLLEGE LONDON (UCL). Information Services Division. UCL Research Information and IT Services Group (RIISG). **Research Data Policy**, 2013, 6 p. Disponível em: https://www.ucl.ac.uk/isd/sites/isd/files/uclresearchdatapolicy_2018.pdf. Acesso em: 03 set. 2020.

UZWYSHYN, R. Research Data Repositories: The What, When, Why and How. **Computers in Libraries: complete Coverage of Library Information Technology**, v. 36, n. 3, p. 18-21, apr. 2016. Disponível em: <https://www.infotoday.com/cilmag/apr16/Uzwyshyn--Research-Data-Repositories.shtml>. Acesso em: 17 jun. 2021.

VAN TUYL, S.; MICHALEK, G. Assessing research data management practices of faculty at Carnegie Mellon University. **Journal of Librarianship and Scholarly Communication**, v. 3, n. 3, p. 1-30, 2015.

VANZ, S. A. S.; STUMPF, I. R. C. Colaboração científica: revisão teórico-conceitual. **Perspectivas em Ciência da Informação**, v. 15, n. 2, p. 42-55, maio./ago. 2010. Disponível em: <https://www.scielo.br/j/pci/a/Fz4q6DhPGhjnhxXmRxLw6Ct/?lang=pt>. Acesso em: 18 jun. 2021.

VICENTE-SÁEZ, R; MARTÍNEZ-FUENTES, C. Open Science now: a systematic literature review for an integrated definition. **Journal of business research**, [s.l.], v. 88, p. 428-436, jul. 2018.

WAZLAWICK, R. S. **Metodologia de pesquisa para Ciência da Computação**. 2. ed. 6ª tir. Rio de Janeiro: Elsevier, 2014. 146 p.

WHITMIRE, A. L.; BOOCK, M.; SUTTON, S. C. Variability in academic research data management practices: implications for data services development from a faculty survey. **Program: electronic library and information systems**, v. 49, n. 4, p. 382-407, 2015.

WILKINSON, M. D. *et al.* The FAIR guiding principles for scientific data management and stewardship. **Scientific Data**, v. 3, 15 mar. 2016. Disponível em: <https://www.nature.com/articles/sdata201618>. Acesso em: 03 set. 2020.

YORK UNIVERSITY. **What is open data?** Toronto. Disponível em: <https://www.library.yorku.ca/web/open/overview/open-data/>. Acesso em: 16 jun. 2021.

YU, H. H. The role of academic libraries in research data service (RDS) provision: opportunities and challenges. **The Electronic Library**, v. 35, n. 4, p. 783-797, 2017.

APÊNDICES

Apêndice A: Formulário de análise para a identificação de práticas de gestão de dados de pesquisa realizadas em diferentes áreas

Dados de identificação	Autor	
	Título do artigo	
	Ano	
	Lugar de publicação	
Caracterização da pesquisa	Objetivo da pesquisa	
	Estratégia de investigação	() Qualitativa () Quantitativa () Mista
	Técnica para a coleta de dados	
	Descritores atribuídos pelo(s) autor(es)	
Identificação de práticas de gestão de dados de pesquisa	Descrição de práticas	() Planejar Descrição:
		() Coletar Descrição:
		() Processar Descrição:
		() Compartilhar Descrição:
		() Preservar Descrição:
		() Descobrir Descrição:
		() Reutilizar Descrição:
		() Monitoramento e controle Descrição:
		() Outra: _____ Descrição:
		Observações

Apêndice B: Roteiro de entrevista

PERGUNTA DE ABERTURA	
1. Considerando a “Ciência Aberta” como um termo guarda-chuva que abrange práticas em acesso aberto que vão desde a produção de dados de pesquisa até o seu reuso, como você percebe essa questão no tópico de pesquisa no qual você atua? Tem se discutido sobre isso?	Vicent-Saez, Martinez-Fuentes (2018); Jomier (2017); Foster Plus.
GESTÃO DE DADOS DE PESQUISA	
2. Como se dá o planejamento dos projetos de pesquisa que você conduz (definição de objetivos da pesquisa, estimativa de recursos, definição de resultados)?	Modelos do ciclo de vida de dados - Universidade de Virginia; DataONE (2013); UK Data Archive; USGS (2013); ICPSR (2012); GFBio; Jisc; DTU; Rüegg <i>et al.</i> (2014).
3. Você já teve alguma experiência na elaboração de um plano de gestão de dados de pesquisa ou qualquer outro instrumento que o auxilie no planejamento da gestão de seus dados? Se sim, como se deu? Se não, por quê?	Modelos do ciclo de vida de dados - Universidade de Virginia; DataONE (2013); UK Data Archive; USGS (2013); ICPSR (2012); GFBio; Jisc; DTU; Rüegg <i>et al.</i> (2014).
4. Geralmente, quais são os procedimentos que você adota para coletar dados em suas pesquisas?	Modelos do ciclo de vida de dados - Universidade de Virginia; DataONE (2013); UK Data Archive; USGS (2013); ICPSR (2012); GFBio; Jisc; DTU; Rüegg <i>et al.</i> (2014).
5. Para efeitos de organização, como você descreve os conjuntos de dados coletados nas pesquisas que você conduz? Há algum padrão de metadados recomendado? Se sim, fale um pouco sobre ele? Se não, por quê?	Modelos do ciclo de vida de dados - Universidade de Virginia; DataONE (2013); UK Data Archive; USGS (2013); ICPSR (2012); GFBio; Jisc; DTU; Rüegg <i>et al.</i> (2014).
6. Qual a importância de documentar os procedimentos adotados em relação aos dados de suas pesquisas, como digitalização, transcrição, tradução, anonimização e interpretação de dados?	Modelos do ciclo de vida de dados - Universidade de Virginia; DataONE (2013); UK Data Archive; USGS (2013); ICPSR (2012); GFBio; Jisc; DTU; Rüegg <i>et al.</i> (2014).
7. Você, ou equipe de algum projeto que participou, já teve alguma experiência de organizar e disponibilizar dados gerados em sua pesquisa? E de forma amplamente aberta? Se sim, como se deu? Se não, por quê?	Modelos do ciclo de vida de dados - Universidade de Virginia; DataONE (2013); UK Data Archive; USGS (2013); ICPSR (2012); GFBio; Jisc; DTU; Rüegg <i>et al.</i> (2014).
8. Na área e no tópico em que você atua, qual a importância de se preservar dados das pesquisas que realiza? Você já teve alguma experiência na elaboração de um plano de preservação de dados de pesquisa a curto ou longo prazo? Se sim, como se deu? Se não, por quê? (Exemplos: cópias de segurança de dados, migração de dados, depósito em repositórios abertos de dados, entre outros)	Modelos do ciclo de vida de dados - Universidade de Virginia; DataONE (2013); UK Data Archive; USGS (2013); ICPSR (2012); GFBio; Jisc; DTU; Rüegg <i>et al.</i> (2014).
9. Na área e no tópico em que você atua, como pesquisadores se posicionam em relação	Modelos do ciclo de vida de dados - Universidade de Virginia; DataONE (2013); UK

ao reuso de dados gerados por outros pesquisadores? (Exemplo: você costuma reutilizar dados de outros pesquisadores? Outros pesquisadores costumam reutilizar dados que você produziu?)

Data Archive; USGS (2013); ICPSR (2012); GFBio; Jisc; DTU; Rüegg *et al.* (2014).

10. Você teria alguma sugestão/recomendação para sua universidade caso a gestão de dados abertos seja institucionalizada?

Wilkinson *et al.* (2016); Patel (2016); Yu (2017).

ABERTURA DE DADOS DE PESQUISA

11. Dados abertos de pesquisa são capazes de desencadear novos ciclos de produção do conhecimento científico quando reutilizados em diferentes áreas do conhecimento. Tendo como pano de fundo conceitos defendidos pela Ciência Aberta, você estaria disposto a compartilhar abertamente os dados de suas pesquisas?

Jomier (2017); Silva (2019); Agência para a Modernização Administrativa - AMA (2016).

12. Quais aspectos estimulariam a abertura de dados de pesquisas realizadas na área em que você atua? E quais aspectos inibiriam a abertura de dados de pesquisas realizadas na área em que você atua?

Jomier (2017); Silva (2019); Agência para a Modernização Administrativa - AMA (2016).

13. Qual a sua opinião sobre a possibilidade de a universidade em que você atua tornar obrigatório o depósito dos dados das pesquisas que você conduz? E além do depósito, a abertura dos dados?

Yu (2017).

14. Qual a sua opinião sobre a possibilidade de agências de fomento tornarem obrigatório o depósito dos dados das pesquisas que você conduz? E além do depósito, a abertura dos dados?

Yu (2017); Federer (2016).

15. Qual a sua opinião sobre a possibilidade de editores científicos tornarem obrigatório o depósito dos dados das pesquisas que você conduz? E além do depósito, a abertura dos dados?

Yu (2017); Federer (2016).

PERGUNTA DE FECHAMENTO

16. Há qualquer outro aspecto relevante que gostaria de considerar a respeito da gestão de dados abertos de pesquisa?

Wilkinson *et al.* (2016); Patel (2016); Sales *et al.* (2019).