

[Liinc em Revista](#), publicada pelo [Instituto Brasileiro de Informação em Ciência e Tecnologia](#), é licenciada sob os termos da [Licença Creative Commons Atribuição 4.0 Internacional – CC BY 4.0](#)



Fonte: <https://revista.ibict.br/liinc/article/view/4763>. Acesso em: 3 nov. 2023.

#### Referência

COSTA, Máira Murrieta; CUNHA, Murilo Bastos da. A necessidade de uma política nacional para a gestão de dados de pesquisa no Brasil. **Liinc em Revista**, Rio de Janeiro, v. 15, n. 2, 2019. DOI: <https://doi.org/10.18617/liinc.v15i2.4763>. Disponível em: <https://revista.ibict.br/liinc/article/view/4763>. Acesso em: 3 nov. 2023.



# A necessidade de uma política nacional para a gestão de dados de pesquisa no Brasil<sup>1</sup>

*The need for a national policy for the management of research data in Brazil*

Maíra Murrieta Costa\*

Murilo Bastos da Cunha\*\*

## RESUMO

Este trabalho discute os conceitos e características da ciência produzida no século XXI. Traz considerações sobre a ciberinfraestrutura necessária à colaboração e compartilhamento de dados de pesquisa. Contextualiza a gestão de dados e os motivos que podem causar impacto no compartilhamento de dados, dentre eles pesquisa em temas de soberania nacional, ou mesmo pesquisas que culminaram como desenvolvimento tecnológico. São apresentados exemplos sobre o processo de compartilhamento de dados em alguns países no exterior. No contexto brasileiro, apresenta e discute o marco legal sobre dados abertos governamentais e científicos. Além disso, apresenta algumas políticas institucionais brasileiras sobre gestão de dados. Ao final do artigo é apresentado um conjunto de diretrizes que pode servir de subsídio para a elaboração de uma política nacional para a gestão de dados de pesquisa.

**Palavras-chave:** Brasil; Gestão de dados de pesquisa; gestão de dados científicos, Informação científica; Política nacional de

## ABSTRACT

This article discusses the concepts and characteristics of science produced in the 21st century. It presents considerations about the necessary cyberinfrastructure for research data collaboration and sharing. It contextualizes data management and the reasons that may impact data sharing, such as research on national sovereignty issues or research that culminated as technological development. Examples are given about the process of data sharing in some countries abroad. In the Brazilian context, it presents and discusses the legal framework on open governmental and scientific data. In addition, it presents some Brazilian institutional policies on data management. At the end presents a set of guidelines that can subsidy the creation of a national policy for research data management.

**Keywords:** Brazil; National information policy; Research data management; Data scholarship, Scientific information.

<sup>1</sup> O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) por meio do Programa de Doutorado Sanduíche no Exterior (PDSE)/ Ciência sem Fronteiras no ano de 2014 – PDSE 99999.004562/2014-03.

\* Doutora em Ciência da Informação pela Universidade de Brasília (UnB). Coordenadora de Inteligência de Negócio e Informação do Ministério da Ciência, Tecnologia, Inovações e Comunicações (MCTIC). Endereço: Esplanada dos Ministérios, Bloco R, CEP 70067-900, Brasília, DF. Telefone: (61) 2027-6267. E-mail: mmurrieta@mctic.gov.br

\*\* Doutor Library da Information Science pela Universidade de Michigan (UMICH). Professor da Faculdade de Ciência da Informação da Universidade de Brasília. Endereço: Campus Universitário, Anexo da Biblioteca Central, CEP: 709199-70, Brasília/DF. Telefone: (61)3107-2635 E-mail: murilobc@unb.br

informação.

## A CIÊNCIA DO SÉCULO XXI E A PRODUÇÃO DE DADOS CIENTÍFICOS

A necessidade da colaboração global de pesquisadores em prol do avanço da pesquisa científica para resolver problemas complexos, tais como o impacto do aquecimento global na produção agrícola, o avanço da medicina na busca da cura do câncer, ou, ainda, os modelos de comportamento elétrico das células cardíacas, dentre outros, mostrou que é preciso se desenvolver uma infraestrutura tecnológica que possibilite aos pesquisadores acessar recursos armazenados em diferentes lugares, por meio de sites, provendo um ambiente de pesquisa robusto e útil, onde diferentes grupos de cientistas possam combinar suas atividades de pesquisa (HEY, TREFETHEN, 2003; GRAY, 2007).

Esse movimento, presente na pesquisa científica do Século XXI, altamente internacionalizada, foi inicialmente denominado como *e-science*, cuja característica é a pesquisa colaborativa, produzida por uma equipe multidisciplinar, que coleta uma grande quantidade de dados, em diferentes lugares. Tais características impulsionaram o chamado dilúvio de dados que precisam ser gerenciados para viabilizar sua preservação e posterior recuperação. Já a infraestrutura tecnológica utilizada para suportar essas pesquisas científicas é denominada ciberinfraestrutura (do inglês, *cyberinfrastructure*).

Essa nova forma de fazer ciência tem colocado em evidência a importância da tecnologia da informação e da gestão de dados de pesquisa (GDP), uma vez que os dados têm sido produzidos em um volume jamais visto. Hey e Trefethen (2003), Gray (2007), bem como Mayer-Schönberger e Cukier (2013) argumentam que os desafios tecnológicos incluem a necessidade de melhor captar, analisar, modelar, visualizar e preservar as informações científicas, o que torna os sistemas de computação vitais para o moderno ambiente de pesquisa. Por outro lado, quando se pensa em acessar os dados produzidos surge a necessidade de se criar mecanismos para organizá-los (catalogar, classificar, indexar, identificar sua origem, sua tipologia dentre outros metadados) de forma a permitir um gerenciamento adequado que viabilize a otimização de recursos de armazenamento para a preservação de longo prazo.

A literatura internacional revela que as iniciativas para tratar o grande volume de dados de pesquisa, seja no âmbito da infraestrutura tecnológica ou da diretriz para a gestão de dados, começaram entre os anos de 2000 e 2003, quando pesquisadores do Reino Unido e dos Estados Unidos alertaram para a necessidade de uma ciberinfraestrutura que desse suporte à colaboração *online* dos pesquisadores. Em 2001, é criado no Reino Unido o National e-Science Centre. Já nos Estados Unidos, a comunidade acadêmica publica em 2003 o *Relatório Atkins* (ATKINS *et al.*, 2003; JANKOWSKI, 2007) com o objetivo de obter fontes de financiamento na National Science Foundation para desenvolvimento dessa ciberinfraestrutura.

Outros países que apresentam iniciativas para a GDP são o Canadá, a Austrália, a Espanha, a Índia, Portugal e as nações signatárias da "Declaração da OCDE sobre o Acesso aos Dados de Pesquisa do Financiamento Público" de 2014.

Enquanto o conhecimento pode ser considerado o motor do desenvolvimento científico e tecnológico, os dados oriundos das pesquisas são o combustível que coloca esse motor em funcionamento. Por meio dessa metáfora entende-se a relevância do tema deste artigo para a sociedade brasileira. Nessa perspectiva, este

trabalho contextualiza a GDP no Brasil e tem como objetivo geral apresentar um conjunto de diretrizes necessárias à elaboração de uma política nacional para a gestão de dados.

O artigo é baseado nos resultados de uma pesquisa exploratória, realizada por Costa (2017), que priorizou o caráter qualitativo na coleta de seus dados como centro do processo de análise do problema de pesquisa. A abordagem de investigação qualitativa utilizada foi a *Grounded Theory* (Teoria Fundamentada em Dados) por esta permitir desenvolver-se uma teoria a partir dos dados sistematicamente recolhidos e analisados, bem como fazer descrições muito úteis (CHARMAZ, 2009; STRAUSS; CORBIN, 2008).

Os tópicos 3, 4 e 5 deste artigo apresentam uma síntese qualitativa<sup>2</sup> da análise dos dados da pesquisa e um *framework* com itens considerados de relevância para a elaboração de um conjunto de diretrizes que venham a servir de elementos norteadores para a elaboração de uma política para a gestão de dados científicos no Brasil.

## **SÍNTESE DO MARCO LEGAL DE ACESSO A INFORMAÇÃO E DADOS DE PESQUISA NO BRASIL**

No Brasil, o acesso à informação e a abertura de dados governamentais possui como primeiro marco fundamental a Constituição Federal Brasileira de 1998, que assegura ao indivíduo o acesso à informação e o direito de receber dos órgãos públicos aquelas informações de seu interesse ou de interesse coletivo ou geral, conforme inciso XXXIII, Art. 5, Capítulo I. A Carta Magna em seu § 2º, do Art. 216 – Seção II dispõe que:

Cabe à administração pública, na forma da lei, a gestão da documentação governamental e as providências para franquear sua consulta a quantos dela necessitem.

O acesso a registros administrativos e informações sobre atos de governo é introduzido no ordenamento jurídico brasileiro por meio da Emenda Constitucional nº 19, de 04 de junho de 1998. Os dispositivos constitucionais acima mencionados foram invocados na Mensagem Presidencial 316/2009, de 15 de janeiro de 2009, que submeteu à deliberação do Congresso Nacional o texto do Projeto de Lei que regula o acesso a informações e que posteriormente deu origem a Lei nº 12.527 de 18 de novembro de 2011, conhecida como Lei de Acesso à Informação (LAI), que em seu Art. 5º dispõe:

É dever do Estado garantir o direito de acesso à informação, que será franqueada, mediante procedimentos objetivos e ágeis, de forma transparente, clara e em linguagem de fácil compreensão.

Para fins de entendimento da abertura de dados de pesquisa científica, merecer ser ressaltado que a LAI define o que é informação em seu Inciso I, Art. 4º como:

Dados, processados ou não, que podem ser utilizados para produção e transmissão de conhecimento, contidos em qualquer meio, suporte ou formato (grifo nosso).

---

<sup>2</sup> Uma análise integrada das respostas concedidas pelo instrumento de coleta de dados aplicado aos doutores envolvidos com o tema (n=40), bem como pelo instrumento de coleta de dados aplicado aos funcionários de agências de fomento e fundações de amparo à pesquisa (n=22).

Sobre a definição de informação contida na LAI, merece ser comentado que ela é inclusiva (o que queremos dizer e que ela não se remete a pirâmide informacional: dado, informação e conhecimento) e não faz distinção entre o dado que promove a transparência e permite um controle social da administração pública (bases de dados governamentais) dos dados oriundos de pesquisa científica.

Ao mesmo tempo em que a LAI era produzida no Congresso Nacional, o Poder Executivo Federal trabalhava na elaboração de um decreto presidencial para dados abertos. Assim, dois meses antes de ser promulgada a LAI, foi publicado o Decreto de 15 de dezembro de 2011, que institui o Plano de Ação Nacional sobre Governo aberto e dá outras providências.

No âmbito internacional, o Brasil faz parte da Parceria Governo Aberto – Open Government Partnership<sup>3</sup> (OGP) desde a sua criação, ocorrida em 20 de setembro de 2011. O Brasil foi representado pela Controladoria-Geral da União como um dos oito<sup>4</sup> países fundadores da OGP que assinaram a Declaração de Governo Aberto<sup>5</sup> e apresentaram seus Planos de Ação.

A partir da publicação da LAI, da participação do Brasil na OGP, surge uma série de dispositivos legais que começam a formar um marco legal para o acesso à informação, à dados governamentais e à dados de pesquisa científica, conforme ilustra a Figura 01.

**Figura 1 – Marco Legal sobre o acesso a informação, a dados governamentais e de pesquisa científica**



Fonte: Costa (2018)

<sup>3</sup> <https://www.opengovpartnership.org/>

<sup>4</sup> Países Fundadores da OGP: África do Sul, Brasil, Estados Unidos, Filipinas, Indonésia, México, Noruega e Reino Unido.

<sup>5</sup> <http://governoaberto.cgu.gov.br/central-de-conteudo/documentos/arquivos/declaracao-governo-aberto.pdf>

## A GESTÃO DE DADOS DE PESQUISA

A pesquisa é um processo de investigação, realizado em qualquer área do conhecimento, que tem como objetivo descobrir as relações existentes entre os aspectos que envolvem os fatos, os fenômenos, as situações ou as coisas. Durante esse processo de investigação o pesquisador vai coletar subsídios para descrever fenômenos, explicar suas relações, validar teorias, dentre outros. É durante esse processo de coleta que o pesquisador cria os registros de sua pesquisa.

Assim, parece haver um consenso na literatura de que dado de pesquisa é a fonte primária, é o primeiro registro do pesquisador quanto à descrição do fenômeno pesquisado. No entanto, Borgman (2015) considera que a definição de dado não é algo trivial, pois a forma de criá-lo, utilizá-lo e compreendê-lo vai variar conforme a comunidade acadêmica à qual o pesquisador pertença. Nas palavras de Sayão e Sales (2015, p. 7), “a noção de dados pode variar consideravelmente entre pesquisadores e, ainda mais, entre áreas do conhecimento”.

O fato é que a variedade de dados produzidos em diferentes áreas do conhecimento, a colaboração entre pesquisadores em prol da celeridade do avanço científico, assim como a infraestrutura tecnológica disponível para que o pesquisador colete um volume de dados jamais visto foram fatores que impulsionaram a necessidade de gerenciar esses dados para permitir seu compartilhamento.

Surgem, a partir da necessidade de GDP, propostas de classificação do dado quanto a sua natureza, a sua origem e o seu nível de reprodutibilidade, bem como nível de processamento (NASA, 2016; BORGMAN, 2015, SIMBERLOF *et al*, 2005).

A GDP envolve ações para administrar o dado, logo subentende-se que compreende as atividades de organização, tratamento técnico, armazenamento/ arquivamento e preservação de longo prazo com o objetivo de deixar o dado disponível para acesso, viabilizando assim seu uso, reuso e seu compartilhamento sempre que necessário.

Entende-se que o principal motivo para efetuar a GDP é viabilizar a utilização e o compartilhamento do mesmo. Nesse sentido, Strasser (2015) argumenta que quando os dados são bem geridos, é mais fácil utilizá-los e viabilizar suas reutilizações, o que se traduz em uma maior colaboração para pesquisadores e um máximo de retorno do investimento para as agências de fomento.

Interessante observar que a literatura revela que a menor dificuldade para se disponibilizar os dados de pesquisa é a questão referente à infraestrutura tecnológica, afinal, a capacidade de *hardware* está cada vez maior, assim como a velocidade de processamento; sendo essa é uma das características do *big data*. Em contrapartida, questões inerentes ao aspecto político e humano são as relacionadas pelos autores como as que geram dificuldades para se disponibilizar o dado. Hey e Thefethen (2002) sintetizam essa questão ao afirmarem que o sucesso para os projetos de *e-Science* não envolvem apenas questões técnicas de infraestrutura tecnológica tais como escalabilidade, confiabilidade, interoperabilidade, tolerância a falhas, gerenciamento de recursos, desempenho e segurança. É preciso atenção para questões inerentes às pessoas envolvidas nos projetos tais como a vontade de trabalhar de forma colaborativa, aceitando o compartilhamento de recursos e dados.

Para gerenciar o dado de pesquisa é necessário mapear um conjunto de metadados técnicos que permitirão classificar, indexar e armazenar os dados, bem como estabelecer sua temporalidade. Além disso, é necessário estabelecer diretrizes para questões legais e éticas de acesso, uso, reuso e compartilhamento. Por fim, em se

tratando de um grande volume de dados *online*, é imperativo identificar as questões de infraestrutura tecnológica para arquivamento e acesso.

De acordo com Costa<sup>6</sup> (2017) é importante identificar todos os envolvidos na pesquisa (líder, pesquisador sênior, pesquisador júnior, demais colaboradores etc.), identificar o tipo de dado coletado, o período de coleta do dado, sua origem (dado primário ou compartilhado), quais os instrumentos utilizados para coletar os dados, os *softwares* necessários para a leitura do dado e o autor do dado. No âmbito das questões legais e éticas, é preciso sinalizar quais as medidas de proteção ao acesso e uso do dado, quais as regras de uso, reuso e compartilhamento dos dados, qual o limite temporal para a reutilização do dado, quais as regras de citação do dado. A Figura 2 sintetiza um conjunto de reflexões sobre o que é necessário para viabilizar a gestão de dados científicos.

Figura 2 – Elementos necessários para a gestão de dados científicos.



Fonte: Costa (2017)

No que diz respeito à citação do dado, Borgman (2015, p. 251-257) explica que a autoria pode, ou não, ser atribuída a quem descreveu a narrativa, a quem coletou os dados, a quem compilou a bibliografia, a quem analisou os dados, ou a quem construiu os instrumentos de coleta dos dados. Por outro lado, relata que os dados raramente são citados. Para a autora, determinar o que constitui uma citação de dados ou o uso em si dos dados, é complexo. Estudos sobre citação de dados indicam que apenas uma pequena porcentagem de artigos inclui a citação de dados nas referências bibliográficas ou notas de rodapé. Apesar disso, há indícios de que o número de citações vêm aumentando em anos mais recentes. Ela esclarece que algumas áreas do conhecimento publicam *papers* de dados e sobre os instrumentos para dar o devido crédito para as contribuições específicas. Em outros casos, autores citam os próprios *papers* para se remeter aos dados contidos nele.

<sup>6</sup> Costa (2017) com fundamento em DCC (201), NASA (2016), DataONE (2016); Borgman (2015); Sayão, Sales (2015), Shearer (2015), Fox, Harris, 2013; Simberlof et al. (2005).

Para exemplificar a complexidade da autoria e citação de dados, Borgman (2015) retoma o primeiro artigo sobre o *Bóson de Higgs* e explica que a instituição CERN identificou os autores em um Atlas de Colaboração e os listou em uma relação onde constaram 2.932 nomes! Essa atitude permaneceu por um determinado período, de forma a garantir que aqueles que contribuíram com os estágios iniciais da pesquisa, pudessem receber os devidos créditos pelas descobertas. Em consequência disso, há casos de autoria póstuma.

## POLÍTICAS PARA GESTÃO DE DADOS CIENTÍFICOS

Para Corrêa (2016), é fato que a Ciência do Século XXI produz um volume de dados na ordem de *terabytes* e *zetabytes*. Da mesma forma, segundo as ideias de Gray (2007), as características dessa nova ciência são a colaboração de pesquisadores por meio de uma infraestrutura tecnológica. Isso posto, ao se propor desenvolver políticas para a GDP é preciso compreender que a colaboração de pesquisadores não se resume à participação em uma reunião por videoconferência, ou mesmo ao compartilhamento de infraestrutura de pesquisa (laboratórios, banco de dados, sistemas de visualização de dados dentre outros).

A colaboração que a GDP prega é o compartilhamento de dados científicos brutos, que dado ao seu volume são processados em um *grid* computacional. A literatura internacional, conforme observado por Tripathi *et al.* (2017), revela que já há órgãos de financiamento de pesquisa e editoras internacionais obrigando o pesquisador a disponibilizar seus dados brutos, seja para permitir a reutilização de dados, ou mesmo para validar a pesquisa feita. A respeito do assunto, é importante ressaltar que o periódico *Nature*, por meio de sua política editorial, passou a exigir, no primeiro semestre de 2013, que os artigos, quando submetidos, fossem acompanhados dos dados brutos, materiais e métodos intrínsecos ao mesmo, logo, indo ao encontro do conceito de publicação ampliada.

Nesse cenário, emerge, de forma intensa, a necessidade de estruturar o gerenciamento do ciclo de vida desses dados. Porém, é preciso levar em consideração que, além do aspecto técnico de GDP, há condutas a serem trabalhadas. Nesse contexto, Costa (2017) argumenta que à medida que a tecnologia da informação avança e as barreiras do livre acesso à informação científica e tecnológica se rompem, o cientista se depara com dúvidas complexas, que envolvem aspectos éticos e políticos. É preciso um norteamento para o cientista para questões tais como – Qual o limite e as regras para a reutilização de dados científicos? Qual o prazo de embargo dos dados de pesquisa que culminou com o desenvolvimento tecnológico? O compartilhamento de dados pode ocorrer apenas com pesquisadores brasileiros (soberania nacional), ou pode ocorrer com pesquisadores estrangeiros?

Quando a instituição de pesquisa não possui um direcionamento para o cientista sobre o que pode e o que não pode ser feito com o dado, as dúvidas não são poucas. Em alguns casos, a diretriz institucional chega a ser insuficiente dada a complexidade da questão, como por exemplo – se o pesquisador receber fomento de agências internacionais e elas exigirem o compartilhamento de dados como contrapartida do apoio, qual a orientação política do Brasil quanto ao caso?

Grande parte das indagações acima precisam ser analisadas em diversos níveis da esfera governamental de um país. Assim, parece coerente afirmar que um conjunto de respostas às indagações precisa ser elaborado por meio de uma política pública para a GDP.

Dessa forma, as políticas de GDP não devem se resumir a questões de infraestrutura tecnológica e de definição de padrões de metadados que viabilizem a sua interoperabilidade. Autores como Hey e Trefethen (2002) e Shearer (2015) defendem que essas políticas também precisam abordar questões de normas de compartilhamento e reutilização de dados, mecanismos de fomento ao compartilhamento de dados, bem como de reconhecimento aos cientistas que disponibilizam seus dados dentre outros aspectos.

No que diz respeito ao compartilhamento de dados, Hey e Trefethen (2002) argumentam que o compartilhar dados é uma excelente forma de otimizar os recursos de financiamento investidos na pesquisa. No entanto, na visão dos autores, a motivação individual que o cientista tem para compartilhar seu dado não é tão clara. Nesse sentido, eles conjecturam que talvez as agências de fomento precisem acrescentar algum incentivo para encorajar o compartilhamento. No Brasil, Costa (2017) identificou que os pesquisadores gostariam que a CAPES, em seu sistema de avaliação dos Programas de Pós-Graduação, atribuísse nota a quem criou o dado, a exemplo do conceito de autoria de documento<sup>7</sup>, ou seja, quem criou o artigo. Nesse sentido, pesquisadores manifestaram interesse de que a plataforma Lattes permitisse inserir informação sobre dados brutos (autoria, tipologia, origem, formato, instrumento de coleta e período de coleta). Na visão deles, o fato da Plataforma Lattes permitir que o pesquisador inclua informações sobre o dado coletado já é uma forma de reconhecimento. Ainda sobre o assunto, Corrêa (2016) é enfático ao afirmar que poucos pesquisadores realmente se preocupam com o registro e a preservação de seus dados; de um modo geral eles mantêm apenas aquilo que necessitam.

Tripathi *et al.* (2017) defendem que os dados qualitativos gerados nas Ciências Sociais também devem ser organizados de forma a permitir o compartilhamento e a reutilização. Porém, na percepção dos autores, esses pesquisadores (Ciências Sociais) estão relutantes em compartilhar dados em função do risco da violação de privacidade do indivíduo que participou da pesquisa. No Brasil, a mesma situação foi verificada por Costa (2017) entre pesquisadores das Ciências Sociais. Os dados do seu trabalho revelaram que apesar de se conseguir anonimizar o dado referente a um indivíduo que participou da pesquisa, há outras questões que preocupam os pesquisadores, como, por exemplo: um conjunto de dados de indivíduos com AIDS quando compartilhados, permite que se construa estereótipo, por vezes preconceituoso, do indivíduo com AIDS no Brasil.

No que diz respeito à GDP<sup>8</sup>, Shearer (2015) argumenta que inúmeros governos e agências de fomento começam a elaborar políticas públicas relacionadas com a gestão desses dados (por vezes nomeada de gestão de dados de pesquisa, do inglês *research data management- RDM*). Na perspectiva da autora citada, geralmente essas políticas tem como objetivo fomentar a colaboração entre pesquisadores de forma a

---

<sup>7</sup> Em face do exposto, retoma-se aqui a complexa discussão apresentada por Borgman (2015) sobre autoria e citação do dado. Ressalta-se que este artigo não discute o tema (pois esse tema, por si só, dá origem a outro artigo), apenas apresenta de forma resumida a complexidade na percepção de Borgman (2015).

<sup>8</sup> Ainda não há um consenso na literatura quanto ao uso da expressão *dados científicos* ou *dados de pesquisa*. Os autores Hey e Hey (2006), Bell (2011), Rodrigues *et al.* (2010) e Costa (2017) utilizam dados científicos. Por outro lado, Borgman (2015), Sales (2014), Sayão e Sales (2014) utilizam o termo dados de pesquisa (*data scholarship*). Há indícios de que o termo “dados de pesquisa” tende a se consolidar como o escolhido pela área.

acelerar o processo de investigação, bem como motivar a reutilização de dados de forma a ampliar a eficiência da pesquisa e facilitar a prestação de contas.

Para Shearer (2015), as políticas têm um escopo comum, normalmente relacionado aos padrões de metadados e às questões de compartilhamento *versus* retenção de dados (dados que não podem ser abertos e compartilhados em função de questões de propriedade intelectual). Mas o fato é que apesar do escopo comum, as políticas de gestão de dados precisam estar alinhadas à missão institucional (e às vezes a um acordo internacional, como, por exemplo, a Declaração da OCDE), devendo dar ênfase a alguns requisitos sobre outros.

Assim, uma política baseada no princípio do compartilhamento de dados provavelmente se concentrará nas práticas-chave necessárias para fornecer acesso aos dados, ao passo que uma política baseada na administração de dados se concentrará nas funções e responsabilidades envolvidas no gerenciamento de dados. Com fundamento no exposto, Shearer (2015, p. 8) apresenta os elementos comuns em uma política de GDP, conforme descrito no Quadro 1.

**Quadro 1 – Elementos comuns em uma política de GDP**

<b>Requisitos da Política</b>	
Qualidade e padrões de dados	Os investigadores são obrigados a aderir aos padrões internacionais para permitir o acesso e a reutilização. A documentação de dados e os metadados devem acompanhar os dados da coleção para que os mesmos sejam compreensíveis por outros pesquisadores
Acesso e compartilhamento de dados	Os investigadores são obrigados a disponibilizar os dados para serem partilhados (normalmente após a publicação dos resultados ou pouco depois, embora algumas agências autorizem períodos de embargo). Requisitos para o depósito de metadados em um catálogo local ou nacional.
Retenção e preservação de dados	Os dados devem ser mantidos por um período mínimo. Sempre que possível, os investigadores devem depositar os seus dados num arquivo de longo prazo para garantir a preservação dos seus dados.
Planos de gestão de dados	As propostas de pesquisa devem incluir um plano de gestão de dados.
<b>Disposições comuns às políticas</b>	
Privacidade	Os direitos e a privacidade dos indivíduos que participam da pesquisa devem ser protegidos em todos os momentos. Assim, os dados disponibilizados para uso mais amplo devem estar livres de identificadores que permitam ligações a participantes individuais da pesquisa e variáveis que podem levar à divulgação dedutiva da identidade de sujeitos individuais.
Conhecimento tradicional	No que se refere aos conhecimentos locais e tradicionais, os direitos dos detentores de conhecimentos não devem ser comprometidos.

Dados de natureza sensível	Quando a liberação de dados pode causar danos, aspectos específicos dos dados podem precisar ser protegidos (por exemplo, localização de ninhos de aves ameaçadas de extinção e localização de santuários ecológicos - locais sagrados)
Propriedade intelectual e dos dados	Poderá ser necessário, por vezes, atrasar a publicação por um curto período para permitir a elaboração do pedido.
<b>Outros aspectos</b>	
Princípios	As políticas de dados aderem a um conjunto de princípios gerais que articulam seu valor.
Âmbito / Cobertura da Política	Descreva o escopo dos dados cobertos pela política.
Funções e Responsabilidades	A política identifica as várias partes responsáveis pela gestão dos dados nas diferentes fases do ciclo de vida.
Acompanhamento e execução	Os meios pelos quais as políticas serão monitoradas ou aplicadas são descritos na política.

Fonte: Shearer (2015, p. 8).

Já na perspectiva do Digital Curation Center (DCC), o *checklist* para planejar uma gestão de dados é composto por sete itens, conforme descrito no Quadro 2.

#### Quadro 2 – Checklist para a gestão de dados de pesquisa.

Coleção de dados	Quais dados serão coletados ou criados? Como os dados serão coletados ou criados?
Documentação e Metadados	Que padrões, documentação e metadados irão acompanhar os dados?
Ética e Compliance Legal	Como serão tratadas as questões éticas? Como serão gerenciados os direitos autorais e direitos de propriedade intelectual (DPI)?
Armazenamento e Backup	Como os dados serão armazenados e apoiados durante a pesquisa? Como o acesso e a segurança serão gerenciados?
Retenção e Preservação	Que dados devem ser conservados e / ou preservados? Qual é o plano de preservação de longo prazo para os dados?
Compartilhamento de dados	Como os dados serão compartilhados? São necessárias restrições na partilha de dados?
Responsabilidades e recursos	Quem será responsável pelo gerenciamento de dados? Que recursos serão necessários para entregar o plano de gerenciamento de dados?

Fonte: Digital Curation Center (2013).

De acordo com Shearer (2015, p. 9) o Reino Unido emitiu em 2011 um conjunto de princípios comuns sobre política de dados intitulado “*Common Principles on Data Policy*”. Esse conjunto de princípios exige que os “dados sejam disponibilizados abertamente com o menor número possível de restrições”. A partir dessa publicação, foram implementadas várias políticas de acesso a dados de pesquisa, a exemplo do Wellcome Trust, que financia pesquisa biomédica.

Nos Estados Unidos, a orientação da National Science Foundation (NSF)<sup>9</sup> é clara quanto à ênfase no compartilhamento:

espera-se que os pesquisadores compartilhem com outros pesquisadores, não apenas com o custo incremental e dentro de um prazo razoável, os principais dados, amostras, coleções físicas e outros materiais de apoio criados ou reunidos no decorrer do trabalho sob financiamento da NSF. Espera-se que os beneficiários incentivem e facilitem essa partilha. As informações privilegiadas ou confidenciais devem ser divulgadas somente em uma forma que proteja a privacidade de indivíduos e sujeitos envolvidos. Os ajustes gerais e, quando essenciais, as exceções a essa expectativa de compartilhamento podem ser especificadas pelo financiamento do Programa NSF ou Divisão / Escritório para um campo ou disciplina particular para salvaguardar os direitos de indivíduos e assuntos, a validade dos resultados ou a integridade das coleções ou para acomodar o interesse legítimo dos investigadores. (NATIONAL SCIENCE FOUNDATION, 2017)

As orientações sobre o compartilhamento de dados devem ser mais específicas e estar em sintonia com a política de desenvolvimento científico e tecnológico do país. Por exemplo, o Brasil permitirá o compartilhamento de dados com pesquisadores de países com os quais não possui relação diplomática, comercial ou acordo de cooperação?

A respeito do assunto, a Índia promove o compartilhamento e a reutilização de dados de pesquisa em Ciências Sociais. Porém, em seu documento ICSSR Data Service possui diretrizes precisas para o compartilhamento de dados, o que, por sua vez, é permitido entre estudantes de graduação, pós-graduação, pesquisadores e membros de institutos de pesquisa do país. Instituições de pesquisa que estejam trabalhando sob o governo da Índia também têm acesso aos dados. Já os pesquisadores estrangeiros<sup>10</sup> podem (grifo nosso) ter acesso aos dados desde que as instituições às quais o pesquisador esteja vinculado tenham acordo bilateral sobre partilha de recursos com o ICSSR e que condições semelhantes sejam retribuídas por essas instituições em relação aos recursos detidos por eles (ICSSR Data Service, 2018).

A questão de compartilhamento de dados é tão complexa que a Biblioteca da Universidade da Califórnia disponibilizou um guia<sup>11</sup> para o pesquisador que permite rastrear, comparar e entender as políticas federais de compartilhamento de dados

---

<sup>9</sup> Disponível em [https://www.nsf.gov/pubs/policydocs/pappguide/nsf13001/aag\\_6.jsp#VID4](https://www.nsf.gov/pubs/policydocs/pappguide/nsf13001/aag_6.jsp#VID4)

<sup>10</sup> Definidos na política do ICSSR como pesquisadores independentes, ONGs baseadas em pesquisa, estudantes, pesquisadores, cientistas e membros do corpo docente de instituições estrangeiras.

<sup>11</sup> Disponível em < <http://datasharing.sparcopen.org/compare?ids=&compare=data> >

por instituição de fomento nos EUA. Interessante observar que as pesquisas com fomento do Departamento de Defesa dos Estados Unidos possuem restrições:

qualquer informação oficial do DoD<sup>12</sup> destinada à liberação pública relativa a questões militares, questões de segurança nacional ou assuntos de interesse significativo para o Departamento de Defesa deve ser revisada para apuramento antes da liberação. (UNITED STATES. DoD Directive 5230.09, 2016)

Shearer (2015) argumenta que de acordo com uma visão geral publicada pela Universidade de Bath, as políticas de GDP geralmente cobrem os seguintes elementos: a) tipos de dados abrangidos pela política; b) expectativas de partilha de dados; incluindo acesso e prazos; c) períodos mínimos de retenção de dados; d) utilização de metadados e normas de documentação; e) exceções justificadas à partilha de dados; f) custos associados à gestão de dados que podem ser pagos por meio de subvenções; e g) reconhecimento de criadores de dados.

A partir do momento que o país possui uma diretriz nacional, as instituições de pesquisa adquirem condições de elaborar políticas institucionais alinhadas estrategicamente à diretriz e, da mesma forma, as agências de fomento podem ser mais enfáticas ao cobrarem um plano de gestão de dados do pesquisador. Na percepção de Shearer (2015) os planos de GDP são geralmente necessários no contexto dessas políticas, já que obrigam os investigadores a pensarem sobre como eles irão gerenciar seus dados antes do projeto ter se iniciado, um requisito chave para as boas práticas de gestão.

Por todo o exposto, fica evidente que o Brasil carece de um posicionamento uniformizado para todas as suas instituições de pesquisa e para todos os seus pesquisadores. A literatura revela que as boas práticas de gestão de dados científicos dependerão de múltiplos fatores contribuintes, incluindo incentivos, conhecimentos especializados, serviços e infraestruturas, bem como mecanismos de financiamento adequados. Os países que optaram por investir na gestão e no compartilhamento de dados científicos descobriram que, embora a conformidade total de uma política não possa ser esperada imediatamente, uma diretriz governamental ajuda muito na conscientização da gestão de dados científicos (ICSSR DATA SERVICE, 2018; UNIVERSITY OF CALIFORNIA, 2018; BORGMAN, 2015; SHEARER, 2015).

Ao se considerar que as políticas governamentais têm procurado auxiliar o Brasil a sair da condição de país emergente para integrar o privilegiado rol dos países desenvolvidos, bem como que o acesso à informação científica e tecnológica é essencial ao desenvolvimento econômico de um país, entende-se que uma diretriz nacional para a GDP deve ser uma ação prioritária da política de ciência, tecnologia e inovação. Ao mesmo tempo, é preciso reconhecer que uma política nacional levará alguns anos para ser totalmente implementada em nosso país. Afinal, a dimensão territorial e as profundas diferenças entre as suas regiões dificultam um alinhamento de ações e evidenciam que as ações de gestão de dados de pesquisa devem ser implementadas gradualmente.

---

<sup>12</sup> Sigla para Department of Defense.

## INICIATIVAS DE GESTÃO DE DADOS CIENTÍFICOS NO BRASIL

No Brasil, a problemática dos dados oriundos da gestão de dados ainda é pouco trabalhada. Vaz (2011) faz uma comparação entre o Reino Unido e Brasil. O autor concluiu que, em 2011, no Brasil, poucos cientistas tinham conhecimento ou interesse sobre na GDP, demonstrando o atraso do país nesse cenário.

Sobre o assunto, a percepção de Sales (2014; p. 49) é de que: “os pesquisadores, as instituições acadêmicas e as agências de fomento à pesquisa começam a entender que estes dados, se preservados e bem gerenciados, constituem uma excelente fonte de recursos informacionais que podem ser compartilhados e reutilizados como insumo para novas pesquisas”.

Corroborando a percepção de Sales (2014), o estudo conduzido por Costa (2017) relata que o Brasil já possui algumas iniciativas de tratamento e disponibilização do dado bruto para consulta *online*. São exemplos: o Portal da Biodiversidade, o Repositório de Dados do Programa de Pesquisa de Biodiversidade da Amazônia Ocidental (PPBIO<sup>13</sup>), o Repositório de Dados do Programa de Pesquisas Ecológicas de Longa Duração (PELD<sup>14</sup>), o Portal GEOINFO de infraestrutura de dados espaciais da EMBRAPA e a plataforma CarpeDIEN<sup>15</sup> (Dados e Informações em Engenharia Nuclear) desenvolvida pelo Instituto de Energia Nuclear.

Por outro lado, em termos de política explícita<sup>16</sup> de informação para a gestão de dados de pesquisa, observa-se que algumas instituições de pesquisa, ainda que de forma embrionária, usam sua autonomia para desenvolver políticas locais que atendam aos editais de fomento internacionais, bem como às necessidades de diretrizes quanto ao armazenamento, à preservação e reutilização de dados, a exemplo, cita-se as instituições ICMBIO, Museu Emílio Goeldi e Jardim Botânico do Rio de Janeiro. Além dessas, há políticas desenvolvidas especificamente para alguns programas de biodiversidade, em função de regras impostas ao recebimento de apoio internacional. São exemplos dessa situação a política do Portal Brasileiro da Biodiversidade e a dos programas PELD e PPBIO. Em nível federal, há apenas o Decreto nº 6.666 de 2008, referente à informação geoespacial. A Figura 3 relaciona as políticas de dados científicos vigentes no Brasil em 2017.

---

<sup>13</sup> O PPBio foi criado em 2004 com o objetivo de desenhar uma estratégia de investimento em CT&I que aponte prioridades, integre competências em diversos campos do conhecimento e dissemine informações sobre biodiversidade que possam ser utilizadas para diferentes finalidades.

<sup>14</sup> O PELD, criado em 1999, é uma iniciativa pioneira no sentido de obter informações relevantes para a conservação da biodiversidade e uso sustentável dos recursos naturais dos ecossistemas brasileiros. As informações coletadas no PELD, que incluem longas séries temporais de dados sobre os ecossistemas e sua biota associada.

<sup>15</sup> Disponível em: <<http://carpedien.ien.gov.br/>>. Acesso em: 2 out. 2016.

<sup>16</sup> Herrera (1995) apresenta os conceitos sobre política implícita e política explícita no âmbito das políticas científicas. A primeira é a política oficial, expressa em leis, declarações governamentais, regulamentos e instituições de governo responsáveis pelo planejamento da ciência. Já a política implícita é a que determina o papel da ciência na sociedade, é muito mais difícil de ser identificada porque não tem estrutura formal; essencialmente expressa as demandas científicas e tecnológicas do "projeto nacional" de cada país, é a que está verdadeiramente em ação. Essas políticas (explícita e implícita) não são necessariamente contraditórias e divergentes. Porém, quando o país está em crise, elas normalmente apresentam contradições.

Figura 3 – Políticas de Dados Científicos vigentes no Brasil em 2017.



Fonte: Costa (2018) com fundamento na literatura revisada e coleta de dados por meio de entrevista e questionário.

Em termos de política para a GDP, é possível afirmar que as áreas de ponta no Brasil são a geoespacial, a área de meio ambiente, ecologia e biodiversidade, pois essas possuem uma política explícita, ainda que incipiente, para a gestão de dados. Essas poucas políticas de GDP foram desenvolvidas em função da maturidade internacional da área de pesquisa no tema, bem como em função da necessidade dos pesquisadores brasileiros se alinharem às diretrizes internacionais desse tipo de informação tanto para obterem financiamento internacional, como para compartilharem seus dados em repositórios internacionais – dando visibilidade à sua pesquisa, ou até mesmo para viabilizar a publicação de um artigo em periódico internacional.

A única iniciativa nacional observada é a infraestrutura nacional de dados espaciais (INDE) criada por meio do Decreto nº 6.666 de 27 de novembro de 2008 que “institui, no âmbito do Poder Executivo federal, a Infraestrutura Nacional de Dados Espaciais - INDE, e dá outras providências”.

A consulta à literatura internacional revela que as iniciativas mundiais para tratar a informação geoespacial iniciam na década de 1990. No Brasil, essas iniciativas ganham corpo na década de 2000, e a criação de um marco legal para o país exigiu a articulação de diferentes atores e instituições, em diferentes níveis governamentais. A partir de 2008, por meio da publicação do Decreto da INDE, percebe-se a criação de uma legislação específica para a área, que culminou com a padronização de marcos geodésicos, dentre outras atividades, que enfim culminaram com a criação de padrões de interoperabilidade para esse tipo de informação.

As instituições brasileiras envolvidas na INDE são o Instituto Brasileiro de Geografia e Estatística (IBGE) e a Secretaria de Planejamento e Investimentos Estratégicos do Ministério do Planejamento, Orçamento e Gestão. A Gestão da INDE, por sua vez, é realizada por meio de um Conselho Superior, um Conselho Consultivo, um Comitê Técnico e Grupos de Trabalho.

Em 2017, o Portal Brasileiro de Dados Espaciais disponibilizava para consulta um conjunto de normas referentes a: a) padronização de marcos geodésicos; b) caracterização do Sistema Geodésico Brasileiro; c) parâmetros para transformação de Sistemas Geodésicos; d) recomendações para levantamentos relativos estáticos; e) normas técnicas da cartografia nacional; f) perfil de metadados Geoespaciais do e Brasil – (Perfil MGB); e g) e-PING padrões de interoperabilidade do governo eletrônico. (INDE, 2017)

No âmbito da informação sobre a biodiversidade, o Portal de Biodiversidade teve em seu desenvolvimento o auxílio de pesquisadores da Escola Politécnica da USP, que conseguiram reunir em uma única interface de busca as informações de bancos de dados mantidos pelo ICMBio e pelo Jardim Botânico do Rio de Janeiro. O Portal oferece buscas textuais e geoespaciais, visualização e *download* de registro de ocorrências de espécies. Além disso, “já conta com mais de um milhão de registros (coordenadas geográficas) de espécies, resultantes da integração de nove bases de dados mantidas pelo ICMBio” (BRASIL. ICMBio, 2015).

## DIRETRIZES PARA UMA POLÍTICA NACIONAL DE GESTÃO DE DADOS CIENTÍFICOS

Rosenberg (1982a, 1982b) já havia alertado, na década de 1980, que os avanços da tecnologia de informação, bem como, as novas formas de comunicação precisavam lançar um novo olhar sobre as regras, regulações, leis e políticas que permeiam o mundo da informação. Como, então mensurar agora em épocas de *big data* a dimensão das modificações necessárias a uma política nacional de informação?

É imperioso comentar que o movimento de *big data* impõe a necessidade de haver um repositório de dados tratados, com possibilidade de recuperação e difusão da informação para a utilização em novas pesquisas. Portanto, faz-se necessário que o Brasil implemente uma política de GDP, como parte de uma política nacional de informação. Para ampliar o entendimento sobre o tema, é preciso discutir sobre políticas de acesso à informação de forma explícita, uma vez que as atuais políticas de informação se mostram obsoletas no que diz respeito aos meios legais de reutilização de dados e proteção à privacidade.

No Brasil, a política de informação é discutida, no âmbito da informação científica e tecnológica, por Aguiar (1980), Garcia (1980), Lemos (1986), Gomes (1988), Costa (1991) e Cunha (2005); da informação arquivística, por Jardim (1999, 2008); da informação sobre telecomunicações por Marques e Pinheiro (2011); da informação para a leitura por Silva, Bernardino e Nogueira (2012); e, no âmbito de uma política de informação geral, por Amaral (1991), Silva (1991), Aun (1999), Unger e Freire (2008). Além destes, Gonzáles de Gómez (1987; 2002) traz importantes contribuições teóricas para o cenário brasileiro.

A literatura revela que a política de informação no Brasil mostra-se imbricada com a política de desenvolvimento científico e tecnológico, ocorrida na década de 1950, em função da necessidade do acesso à informação científica e tecnológica. O fato que marca a relação entre as duas políticas é uma das atribuições do CNPq, quando de sua criação, qual seja, “manter relação com instituições nacionais e estrangeiras para intercâmbio de documentação técnico-científica” (FERREIRA, 1999). Essa atribuição culminou com a publicação do Decreto nº 35.124, em 27 de fevereiro de 1954, que criou o Instituto Brasileiro de Bibliografia e Documentação (IBBD), órgão vinculado ao CNPq. Merece lembrar que em 1976 o IBBDD enfrentou transformações que

culminaram com a mudança do seu nome para Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT<sup>17</sup>),

Sobre o assunto, Gomes (1988, p. 105) já defendia a ideia de que “as orientações fundamentais relativas à definição e à implantação de uma política nacional de informação científica e tecnológica já estão implícitas no planejamento econômico e social em geral e, particularmente, na política de desenvolvimento científico e tecnológico”.

É pertinente comentar que Lemos (1986), Tarapanoff (1992) e Cunha (2005) consideram o IBICT o órgão responsável pela execução da política brasileira de informação científica e tecnológica.

A percepção de Costa (2017) é que a elaboração de diretrizes para uma política de GDP deve envolver os diferentes atores de Ciência e Tecnologia em nível altamente estratégico, como, por exemplo, o Ministério da Ciência, Tecnologia, Inovações e Comunicações, o Ministério da Educação, a Comissão de Ciência e Tecnologia do Congresso Nacional, o presidente das respectivas agências de fomento (CAPES, CNPq e FAP). Além destes, também precisam ser envolvidos os institutos de pesquisa do MCTIC, bem como os demais institutos, as universidades federais e estaduais. A participação desses trará para a discussão a necessidade real de pesquisadores, como, por exemplo, os que trabalham com dados espaciais (AEB, INPE), ou dados sobre a biodiversidade (Museu Emílio Goeldi, INPA), ou ainda as necessidades inerentes ao tratamento de dados de energia nuclear (CNEN, IEN). Além disso, a discussão também deve abordar aspectos de infraestrutura computacional para transmissão e armazenamento de dados (RNP, LNCC), infraestrutura para tratar o ciclo documental do dado científico e suas particularidades para armazenamento e preservação de longo prazo (IBICT).

No que diz respeito à elaboração de uma diretriz nacional para a GDP, merece ser ressaltado que o Ministério da Ciência, Tecnologia, Inovações e Comunicações a Portaria nº 6.072 de 23 de novembro de 2018 que instituiu o Grupo de Trabalho para a **elaboração de proposta de Política Nacional para Ciência Aberta**. Em função da alteração da estrutura regimental do MCTIC, a Portaria nº 6.072 foi atualizada para Portaria nº 1.515 de 04 de abril de 2019. O Art. 3º da portaria em questão determina que o Grupo de Trabalho tem *180 dias prorrogável justificadamente*, para a conclusão de seus trabalhos (*grifo nosso*).

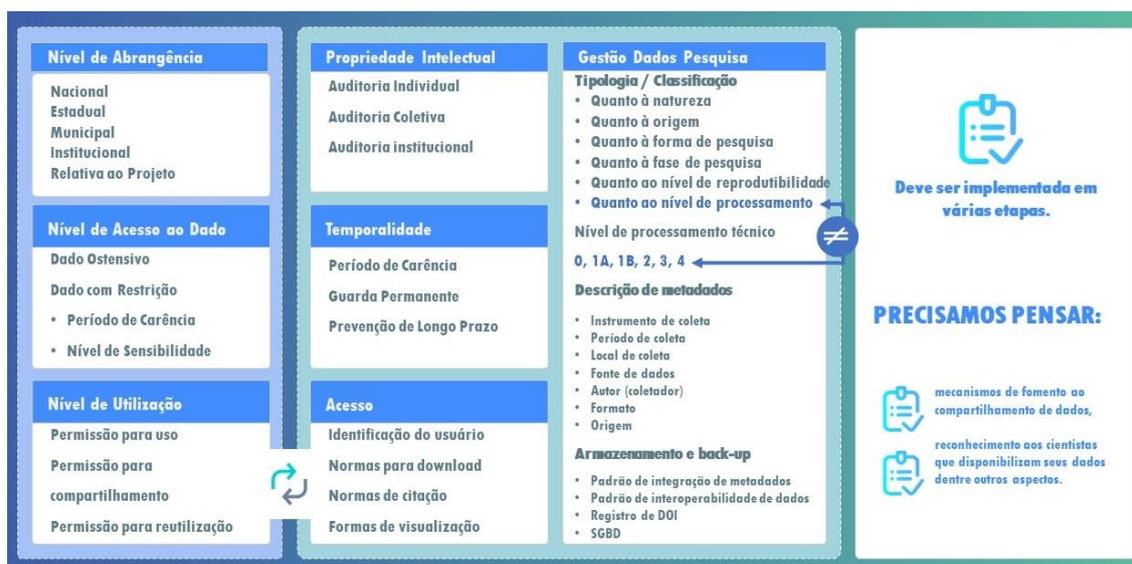
Entre dezembro de 2018 e março de 2019 o GT de Ciência Aberta do MCTIC se reuniu três vezes. Durante as reuniões ficou evidente a necessidade de inclusão de novos *stakeholders* para compor o GT, corroborando a percepção de Costa (2017) no que diz respeito a construção de uma Política Nacional. Portanto, o líder do GT – membro da Secretaria Executiva do MCTIC, solicitou ao Ministro de Estado de Ciência, Tecnologia, Inovações e Comunicações convidasse para integrar o respectivo Grupo a CAPES, a Presidente do Conselho Nacional das Fundações de Amparo à Pesquisa (CONFAP), bem como, a FAPESP, em função do seu pioneirismo no lançamento de editais de pesquisa que exigem do pesquisador o Plano de Gestão de Dados de Pesquisa.

Em termos da elaboração de uma redação para uma política de gestão de dados de pesquisa, recomenda-se que esse documento seja dividido em duas dimensões, a primeira política e a segunda técnica, conforme ilustra a Figura 4.

---

<sup>17</sup> A Resolução nº 20, de 25 de março de 1976 cria o IBICT.

**Figura 4 – Dimensão política e técnica das diretrizes de uma política de dados de pesquisa**



Fonte: Costa (2017) com fundamento em Shearer (2015) e Digital Curation Center (2013).

As diretrizes referentes à dimensão política referem-se à necessidade de que o objetivo da política seja claro, como, por exemplo – garantir o acesso ao conjunto de dados científicos, ou – promover o compartilhamento de dados, dentre outros. Além disso, é importante ser enfática quanto ao seu nível de abrangência e relacioná-la com as formas de financiamento que o pesquisador obteve. Também é necessário que a política se posicione sobre quais dados são ostensivos, quais possuem restrição, seja a restrição em função de período de carência para acesso ao dado, ou da sensibilidade do mesmo. Outrossim, precisa deixar claro quais são as regras para uso, reutilização e compartilhamento dos dados.

Na dimensão técnica, a política precisa estabelecer os critérios de propriedade intelectual do dado, a temporalidade de armazenamento dos mesmos, bem como as regras de acesso. Para tanto, é preciso ter clareza quanto ao processo de gestão de dados de pesquisa que, dentre tantas atividades, compreende a classificação e a descrição de metadados do dado, ao nível de processamento, bem como às formas de armazenamento e *back-up* do dado

A execução da política de armazenamento de dados e preservação, inexoravelmente, deve estar ligada às agências de fomento que apoiam as pesquisas. Para tanto, recomenda-se que as respectivas agências exijam um plano de gestão de dados a qualquer pesquisador que almejar recurso público.

Por outro lado, o papel das universidades não pode ser subestimado. A exemplo do processo de organização e preservação das teses e dissertações produzidas no Brasil, sugere-se que as universidades e os institutos de pesquisa brasileiros dividam a responsabilidade sobre a organização desses dados em parceria com o IBICT. Entende-se que uma instituição sozinha não terá condições de abraçar todo o processo de gestão de dados científicos no Brasil em função da sua complexidade, dimensão territorial, das dificuldades orçamentárias e da escassez de recursos humanos, dentre tantos outros fatores. Nesse cenário, a formação de uma espécie de consórcio para a gestão de dados científicos parece profícua no sentido de aproveitar-se o que cada instituição possui de melhor.

## CONSIDERAÇÕES FINAIS

Na visão de Gray (2007), os formuladores de políticas públicas para a GDP precisam estimular tanto a criação de *softwares* de análise de dados, como o apoio a essas ferramentas. Entende-se aqui que o apoio pressupõe as atividades de sustentação de um sistema já em operação, o que viabiliza o aprimoramento contínuo desses *softwares*.

Borgman (2013) apresentou reflexões sobre questões inerentes ao compartilhamento de dados, como, por exemplo: quais dados devem ser compartilhados, quando os dados podem ser compartilhados, e de que forma os dados podem ser compartilhados. Do ponto de vista do tratamento técnico da informação, a mesma autora tece as seguintes considerações – como atribuir crédito para dados (atribuição) e como fazer referência a dados (citação) de forma que os outros possam identificar, descobrir e recuperá-los?

As experiências no exterior revelam que o sucesso da implementação de uma política interministerial para a GDP dependerá das agências de fomento (CAPES, CNPq, FAP) ao exigir o depósito dos dados coletados em um repositório e ao criarem mecanismos de estímulo ao depósito do dado, como, por exemplo, o peso de reconhecimento da citação do dado. A pesquisa de Costa (2017) também apontou o envolvimento das agências de pesquisa como crucial ao sucesso de uma política de GDP. Nas palavras de um pesquisador entrevistado – “*as universidades só vão se mexer quando uma instância superior disser faça*”.

Ressalta-se que a gestão de dados perpassa inúmeras temáticas e áreas governamentais. Este detalhe, necessariamente, forçará que haja um consenso, entre outros, do Ministério da Ciência, Tecnologia, Comunicações e Inovações (MCTIC), da Educação (MEC), da Saúde (MS), da Indústria e Comércio (MDIC) e da Defesa (MD). É claro que o Legislativo e o Judiciário também deverão estar envolvidos, além das principais agências de fomento (CAPES, CNPq, FINEP, BNDES e as fundações estaduais de amparo à pesquisa).

As questões inerentes à política nacional de informação no Brasil são muitas e complexas. É necessário que o profissional da informação se aprofunde no tema e proponha diretrizes, em conjunto com pesquisadores de cada área do conhecimento, para a construção de um modelo teórico que atenda a um conjunto mínimo de diretrizes para a descrição de dados científicos, o tratamento técnico, a temporalidade de armazenamento (incluindo formas de preservação e regras para descarte), as regras de reutilização e citação do dado original.

Como se pode depreender, para que a política nacional de GDP seja implantada com maior celeridade será importante que ela seja feita em vários momentos. O primeiro deles, a publicação da Portaria nº 1.515 de 04 de abril de 2019 que institui o Grupo de Trabalho para a elaboração de proposta de Política Nacional para Ciência Aberta deixa evidente a emergência do tema e a atenção que este Ministério tem dedicado a ciência aberta, bem como ao acesso aberto de dados científicos.

O Brasil precisa recuperar o seu atraso na gestão dos dados gerados pelas pesquisas financiadas com recursos públicos. É uma questão que poderá refletir no incremento do crescimento nacional, com possível redução nas duplicações de esforços por parte

dos pesquisadores. Além do que, por meio da redução nas duplicações de esforços, é possível otimizar a aplicação de recursos<sup>18</sup> para a pesquisa.

## REFERÊNCIAS

AGUIAR, Afrânio Carvalho. Coordenação de uma rede nacional de informação em ciência e tecnologia: um plano prioritário do IBICT. **Ciência da Informação**, Brasília, v. 9, n. 1/2, p. 83-88, 1980.

AMARAL, Ana Maria Barros Mais do. O cenário da política nacional de informação no Brasil. **Informação & Sociedade**, João Pessoa, v. 1, n. 1, p. 47-53, jan./dez. 1991.

ATKINS, Daniel E. *et al.* **Revolutionizing science and engineering through cyberinfrastructure**: report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure. Washington, DC, January 2003. Disponível em: <<http://www.nsf.gov/cise/sci/reports/atkins.pdf>>. Acesso em: 10 abr. 2015.

AUN, Marta Pinheiro. A construção de políticas nacional e supranacional de informação: desafio para os Estados nacionais e blocos regionais. **Ciência da Informação**, v. 28, n. 2, 1999.

BELL, Gordon. Prefácio. In: HEY, Tony; TANSLEY, Stewart; TOLLE, Kristin (Org.). **O quarto paradigma**: descobertas científicas na era da e-Science. São Paulo: Oficina de Textos, 2011. p. 11-15.

BORGMAN, Christine L. Why are the attribution and citation of scientific data important? In: UHLIR, P. E. **For attribution**: developing data attribution and citation practices and standards: summary of an international workshop. Washington: The National Academic Press, 2013.

BORGMAN, Christine L. Big data, little data, no data: scholarship in the networked world. Cambridge: MIT Press, 2015. 416 p.

BRASIL. Decreto nº 6.666, institui, no âmbito do Poder Executivo Federal, a infraestrutura nacional de dados espaciais - INDE, e dá outras providências. **Diário Oficial [da] República Federativa do Brasil**. Poder Executivo, Brasília, DF, 28 novembro 2008. Seção 1.

Brasil. Decreto nº 35.124, de 27 de fevereiro de 1954. Cria o Instituto Brasileiro de Bibliografia e Documentação, nos termos da Lei nº 1.310, de 15 de janeiro de 1951. **Diário Oficial [da] República Federativa do Brasil**. Poder Executivo, Brasília, DF, 4 MARÇO 1954. Seção 1.

BRASIL. ICMBIO. **MMA e ICMBIO lançam Portal da Biodiversidade**. Brasília, 26 novembro 2015. Disponível em < <http://www.icmbio.gov.br/sisbio/destaques/48-site-ja-tem-mais-de-15-milhao-de-registros-de-ocorrencias-sobre-especies.html>> Acesso em 02 fevereiro 2016.

---

<sup>18</sup> Ressalta-se que a verba disponível para o empenho do Ministério da Ciência, Tecnologia, Inovações e Comunicações caiu 44% na comparação com o recurso inicialmente previsto na LOA, de R\$ 5,049 bilhões, em 2017. Em resumo, os recursos, que já eram praticamente a metade dos cerca de R\$ 10 bilhões registrados em 2013, agora se apresentam como um dos piores da história da ciência, tecnologia e inovação.

Brasil. MCTIC. Portaria nº 6.072 de 23 de novembro de 2018. Institui Grupo de Trabalho para a elaboração de proposta de Política Nacional para Ciência Aberta. **Diário Oficial [da] República Federativa do Brasil. Poder Executivo**, Brasília, DF, 26 NOVEMBRO 2018. Seção 1.

Brasil. MCTIC. Portaria nº 1.515 de 04 de abril de 2019. Altera a Portaria nº 6.072 de 23 de novembro de 2018 que Institui Grupo de Trabalho para a elaboração de proposta de Política Nacional para Ciência Aberta. **Diário Oficial [da] República Federativa do Brasil. Poder Executivo**, Brasília, DF, 5 ABRIL 2019. Seção 1.

CHARMAZ, Kathy. **A construção da teoria fundamentada**: guia prático para análise qualitativa. Porto Alegre: Artmed, 2009.

CORRÊA, Fabiano Couto. **Gestión de datos de investigación**. Barcelona: UOC, 2016. (Colección EPI scholar).

COSTA, Antônio Roberto F. Política nacional de informação científica e tecnológica: necessidade versus realidade. **Informação & Sociedade**, João Pessoa, v. 1, n. 1, p. 30-37, jan./dez. 1991.

COSTA, Maira Murrieta. **Diretrizes para uma política de gestão de dados científicos no Brasil**. 2017. 288 f. Tese (Doutorado em Ciência da Informação). Programa de Pós-Graduação em Ciência da Informação, Universidade de Brasília. Brasília, 2017. Disponível em <<http://repositorio.unb.br/handle/10482/24895>>

COSTA, Maira Murrieta. **Diretrizes para uma política para a gestão de dados científicos no Brasil**. (Apresentação). Seminário de Dados Abertos – Ciclo de Seminários MCTIC 2018. Disponível em <<https://wiki.rnp.br/pages/viewpage.action?pageId=108957257>>

CUNHA, Murilo Bastos da. IBICT: 51 anos. **Ciência da Informação**, Brasília, v. 34, n. 1, p. 7-8, jan./abr. 2005

.DATAONE. **Best practices**. 2016. Disponível em: <<https://www.dataone.org/best-practices>>. Acesso em: 3 jun. 2016.

DIGITAL CURATION CENTER. **Checklist for a Data Management Plan v.4.0**, 2013. Edinburgh: Digital Curation Centre. 2013. Disponível em <<http://www.dcc.ac.uk/resources/data-management-plans>>. Acesso em 20 abril 2017.

FERREIRA, José Rincon. Os 45 anos do IBICT. **Ciência da Informação**, Brasília, Editorial, v. 28, n. 1, 1999.

FOX, Peter; HARRIS, Ray. ICSU and the challenges of data information management for international science. **Data Science Journal**, v. 12, n. 10, February 2013.

GARCIA, Maria Lúcia Andrade. A informação científica e tecnológica no Brasil. **Ciência da Informação**, Brasília, v. 1, n. 1/2, p. 41-81, 1980.

GOMES, Maria Yêda F. S. de Filgueiras. O estado e o processo de implantação de uma política nacional de informação científica e tecnológica no Brasil. **Ciência da Informação**, Brasília, v. 17, n. 2, p. 105-117, jul./dez. 1988.

GONZÁLES DE GÓMEZ, Maria Néilda. Novos cenários políticos para a informação. **Ciência da Informação**, Brasília, v. 31, n. 1, p. 27-40, jan./abr. 2002.

GONZÁLES DE GÓMEZ, Maria Néilda. O papel do conhecimento e da informação nas formações políticas ocidentais. **Ciência da Informação**, Brasília, v. 16, n. 2, p. 157-167, jul./dez. 1987.

GRAY, Jim. e Science: a transformed scientific method. In: NATIONAL RESEARCH COUNCIL (NRC-CSTB), California, 2007. **Talks....** 2007. Disponível em: <[http://research.microsoft.com/en-us/um/people/gray/talks/NRC-CSTB\\_eScience.ppt](http://research.microsoft.com/en-us/um/people/gray/talks/NRC-CSTB_eScience.ppt)>. Acesso em: 30 ago. 2012.

HERRERA, Amílcar O. Los determinantes sociales de la política científica em America Latina: política científica explícita y política implícita. **Red de Revistas Científicas de América Latina, el Caribe, España y Portugal**, v. 2, n. 5, p, 117-131, dic. 1995.

HEY, Tony; HEY, Jessie. E-science and its implications for the library community. **Library Hi-Tech**, v. 24, n. 4, p. 515-528, 2006.

HEY, Tony; TREFETHEN, Anne E. The UK e-science core programme and the grid. **Future Generation Computer Systems**, v. 18, n. 8 p. 1017-1031, 2002.

HEY, Tony; TREFETHEN, Anne. E-science and its implications. **Philosophical Transactions of the Royal Society (A)**, v. 361, p. 1809-1825, June 2003.

INDIAN COUNCIL OF SOCIAL SCIENCE RESEARCH. ICSSR DATA SERVICE. Indian Social Science Repository, **Data deposit and access policy guideline**. 2018. Disponível em <<http://www.icssrdataservice.in/files/ICSSR%20Data%20Service-Policy%20Guidelines.pdf>> Acesso em 15 fevereiro 2018.

INDE. **Sig Brasil**: o portal brasileiro de dados espaciais. Apresentação. 2017. Disponível em: <<http://www.inde.gov.br/a-inde/apresentacao.html>>. Acesso em: 8 março 2017.

JANKOWSKI, Nicholas W. Exploring e-science: an introduction. **Journal of Computer-Mediated Communication**, v. 12, p. 549-562, 2007.

JARDIM, José Maria. Políticas públicas de informação: a (não) construção da política nacional de arquivos públicos e privados no Brasil (1994-2006). In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 2008, São Paulo. **Anais...** São Paulo: ANCIB, 2008.

JARDIM, José Maria. **Transparência e opacidade do Estados no Brasil**: usos e desusos da informação governamental. Niterói, RJ: Editora UFF, 1999.

LEMOS, Antônio Agenor Briquet de. Planejamento e coordenação da informação científica e tecnológica no Brasil. **Ciência da Informação**, Brasília, v. 15, n. 2, p. 107-116, jul./dez. 1986.

MARQUES, Rodrigo Moreno; PINHEIRO, Marta Macedo Kerr. Política de informação nacional e assimetria de informação no setor de telecomunicações brasileira. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 16, n. 1, p. 65-91, jan./mar. 2011.

MAYER-SCHÖNBERGER, Viktor; CUKIER, Kenneth. **Big data**: como extrair volume, variedade, velocidade e valor da avalanche de informação cotidiana. Rio de Janeiro: Elsevier, 2013.

NASA. **Earth Science Data processing levels for EOSDIS data**. 2016. Disponível em: <<http://science.nasa.gov/earth-science/earth-science-data/data-processing-levels-for-eosdis-data-products/>>.

NATIONAL SCIENCE FOUNDATION. Office of Budget Finance & Award Management (BFA). Grant Proposal Guide. **Dissemination and Sharing of Research Results**. 2017. Disponível em <[https://www.nsf.gov/pubs/policydocs/pappguide/nsf13001/aag\\_6.jsp#VID4](https://www.nsf.gov/pubs/policydocs/pappguide/nsf13001/aag_6.jsp#VID4)> Acesso em 15 fevereiro 2018.

RODRIGUES, Eloi ; SARAIVA, Ricardo. **Os repositórios de dados científicos**: estado da arte. Guimarães: Universidade do Minho, 2010. Relatório D-24: Disponível em: <[http://projeto.rcaap.pt/index.php?option=com\\_remository&Itemid=2&func=startdown&id=271&lang=pt](http://projeto.rcaap.pt/index.php?option=com_remository&Itemid=2&func=startdown&id=271&lang=pt)>

ROSENBERG, Victor. National information policies. **Annual Review of Information Science and Technology**, v. 17, p. 3-32, 1982b.

ROSENBERG, Victor. Política de informação nos países em desenvolvimento: o caso do Brasil visto por um americano. **Ciência da Informação**, v. 11, n. 2, p. 37-43, 1982a.

SALES, Luana Farias. **Integração semântica de publicações científicas e dados de pesquisa**: proposta de modelo de publicação ampliada para a área de ciências nucleares. 2014. 268 f. Tese (Programa de Pós-Graduação em Ciência da Informação) – Universidade Federal do Rio de Janeiro, 2014.

SAYÃO, Luís Fernando; SALES, Luana Farias. Dados abertos de pesquisa: ampliando o conceito de acesso livre. **RECIIS. Electronic Journal of Communication Information and Innovation in Health**, v. 8, p. 76-92, 2014.

SAYÃO, Luís Fernando; SALES, Luana Farias. Ciberinfraestrutura de informação para a pesquisa: proposta de integração entre repositório institucional, repositório de dados e CRIS. **Informação & Sociedade**, João Pessoa, v. 25, p. 163-184, 2015.

SHEARER, Kathleen. **Comprehensive brief on research data management policies**. Canada Government, 2015. Disponível em: <<http://docplayer.net/17594465-Comprehensive-brief-on-research-data-management-policies.html>> Acesso em 28 julho 2016.

SILVA, Terezinha Elizabeth da. Política de informação na pós-modernidade: reflexões sobre o caso do Brasil. **Informação & Sociedade**, João Pessoa, v. 1, n. 1, p. 8-13, jan./dez. 1991.

SILVA, Marta Benjamim da; BERNARDINO, Maria Cleide Rodrigues; NOGUEIR, Carine Rodrigues. Políticas públicas para a leitura no Brasil: implicações sobre a leitura infantil. **Ponto de Acesso**, Salvador, v. 6, n. 3, p. 20-46, abr. 2012.

SIMBERLOFF, Daniel *et al.* **Long-lived Digital data collections**: enabling research and education in the 21<sup>st</sup> century. National Science Board, National Science Foundation, 2005. Disponível em: <[https://www.nsf.gov/pubs/2005/nsb0540/nsb0540\\_1.pdf](https://www.nsf.gov/pubs/2005/nsb0540/nsb0540_1.pdf)>

STRASSER, Carly. **Research data management**: a primer publication of the national information standards organization. Baltimore: NISO, 2015. (NISO Premier Series). Disponível em < <http://wiki.lib.sun.ac.za/images/2/24/PrimerRDM-2015-0727.pdf> >

STRAUSS, Anselm; CORBIN, Juliet. **Pesquisa qualitativa**: técnicas e procedimentos para o desenvolvimento da teoria fundamentada. 2. ed. Porto Alegre: Artmed, 2008. 288 p.

TARAPANOFF, Kira. A política científica e tecnológica no Brasil: o papel do IBICT. **Ciência da Informação**, Brasília, v. 21, n. 2, p. 87-166, maio/ago. 1992.

TRIPATHI, Manorama; SHUKLA, Archana; SONKER, Sharad Kumar. Research data management practices in University Libraries: a study. **Journal of Library & Information Technology**, v. 37, n. 6, p. 417-424, November 2017.

UNGER. Roberto José Gervásio; FREIRE, Isa. Maria. Regimes de informação na sociedade da informação: uma contribuição para a gestão de informação. **Revista digital de Biblioteconomia e Ciência da Informação**, Campinas, v. 6, n. 1, p. 87-114,

jan./jun. 2008. Disponível em:  
<<https://periodicos.sbu.unicamp.br/ojs/index.php/rdbci/article/view/2014/2135>>  
Acesso em: 26 abr. 2016.

UNITED STATES. DEPARTMENT OF DEFENSE. **Directive number 5230.09 - Clearance of DoD Information** for Public Release. Incorporating Change 1, Effective March 16, 2016. Disponível em <[https://sparcopen.org/wp-content/uploads/2016/04/DoD\\_Directive\\_523009p.pdf](https://sparcopen.org/wp-content/uploads/2016/04/DoD_Directive_523009p.pdf)> Acesso em 12 fevereiro 2018.

UNIVERSITY OF CALIFORNIA. UCLA Library. **Data Management & Curation Services**. Resources. Data (and Article) Sharing Requirements by Federal Agency. 2018. Disponível em <<http://datasharing.sparcopen.org/>> Acesso em 17 fevereiro 2018.

VAZ, Glauber José. **E-Science na Embrapa**. Campinas: Embrapa Informática Agropecuária, 2011. (Documentos, 117).