

ADRIANA ALVES RODRIGUES
GUILHERME ATAÍDE DIAS
RAFAEL LUÍS GALDINI RAIMUNDO
Autores(as)

GESTÃO DE DADOS DE PESQUISA

Ampliando a geração de
conhecimentos científicos no
contexto da ciência aberta



UNIVERSIDADE FEDERAL DA PARAÍBA

Terezinha Domiciano Dantas Martins

Reitora

Mônica Nóbrega

Vice-Reitora



Pró-Reitoria de Pós-Graduação

Evandro Leite de Souza

Pró-Reitor de Pós-Graduação

Síglia Lima Mendes

Organização Técnica de Seleção - PRPG



Editora UFPB

Geysa Flávia Câmara de Lima Nascimento

Diretora Geral da Editora UFPB

Rildo Coelho

Coordenador de Editoração

GESTÃO DE DADOS DE PESQUISA

Ampliando a geração de
conhecimentos científicos no
contexto da ciência aberta

1ª Edição - 2025

Obra vinculada ao Edital PRPG/UFPB N° 01/2024, financiado pelo Programa de Apoio à Produção Científica (PRÓ-PUBLICAÇÃO DE LIVROS) da Pró-Reitoria de Pós-Graduação, autorizada pelos autores para publicação em formato e-book.

Direitos autorais 2025 - Editora da UFPB



Esta obra é de acesso aberto. É permitida a reprodução parcial ou total desta obra, desde que citada a fonte e autoria e respeitada a Licença Creative Commons indicada.

O CONTEÚDO DESTA PUBLICAÇÃO, SEU TEOR, SUA REVISÃO E SUA NORMALIZAÇÃO SÃO DE INTEIRA RESPONSABILIDADE DOS AUTORES.

OS DIREITOS DE PROPRIEDADE DESTA EDIÇÃO SÃO RESERVADOS À:



Cidade Universitária, Campus I - Prédio da Editora Universitária, s/n
João Pessoa - PB CEP 58.051-970
Site: www.editora.ufpb.br
Instagram: @editoraufpb
E-mail: atendimento@editora.ufpb.br
Fone: (83) 3216.7147

Editora filiada à



ADRIANA ALVES RODRIGUES
GUILHERME ATAÍDE DIAS
RAFAEL LUÍS GALDINI RAIMUNDO
Autores(as)

GESTÃO DE DADOS DE PESQUISA

Ampliando a geração de
conhecimentos científicos no
contexto da ciência aberta

Editora UFPB
João Pessoa
2025

CONSELHO EDITORIAL ESPECIAL - PRPG

(Edital PRPG/UFPB N° 01/2024 - Portaria N° 01/2024 - Reitoria)

Adriana Carla Costa Ribeiro Clementino (Ciências da Saúde - UFPB)
Alexandre Luís Gonzaga (Educação e Letras - UFAC)
Carlos Junior Gontijo Rosa (Educação e Letras - UFAC)
Carlos Xavier de Azevedo Netto (Ciências Sociais Aplicadas - UFPB)
Cleide Vilanova Hanisch (Ciências Sociais Aplicadas - UFPB)
Daniel Germano Maciel (Ciências da Saúde - UFPB)
Eduardo Sérgio Soares Sousa (Ciências Médicas - UFPB)
Giciane Carvalho Vieira (Ciências da Saúde - UFPB)
Italo Roger Ferreira Moreno P. da Silva (Energias Alternativas e Renováveis - UFPB)
José Diego Sales do Nascimento (Ciências da Saúde - UFPB)
José Irialdo Alves Oliveira Silva (Ciências Jurídicas - UFPB)
Manoel Coracy Saboia Dias (Filosofia e Ciências Humanas - UFAC)
Marcelo Rodrigo da Silva (Comunicação Turismo e Artes - UFPB)
Michel Ferreira dos Reis (Educação e Letras - UFAC)
Pedro da Silva de Melo (Educação e Letras - UFAC)

Catálogo na fonte: **Biblioteca Central da Universidade Federal da Paraíba**

R696g Rodrigues, Adriana Alves.
 Gestão de dados de pesquisa : ampliando a geração de conhecimentos científicos no contexto da ciência aberta [recurso eletrônico] / Adriana Alves Rodrigues, Guilherme Ataíde Dias, Rafael Raimundo Galdino. - Dados eletrônicos - João Pessoa : Editora UFPB, 2025.

E-book.

Modo de acesso: <http://www.editora.ufpb.br/sistema/press/>
ISBN : 978-65-5942-300-2

1. Gestão de dados - Plano. 2. Dados de pesquisa. 3. Repositório de dados. 4. Gestão de dados - Política. 5. Governança de dados. 6. Ciclo de vida de dados I. Dias, Guilherme Ataíde. II. Galdino, Rafael Raimundo. III. Título.

UFPB/BC

CDU 004.6

Como citar a publicação no todo (ABNT 6023:2018):

RODRIGUES, Adriana; DIAS, Guilherme; GALDINO, Rafael. **Gestão de dados de pesquisa**: ampliando a geração de conhecimentos científicos no contexto da ciência aberta. João Pessoa: Editora UFPB, 2025. E-book (93 p.). ISBN 978-65-5942-300-2. Disponível em: _____. Acesso em: ____.

PREFÁCIO

Inicialmente registro a imensa satisfação em prefaciar essa obra pela tamanha relevância que o tema apresenta na atualidade. O tema da gestão de dados no âmbito da ciência aberta tem sido divisor de águas no avanço da ciência e acaba por envolver desafios e oportunidades no seu processo de amadurecimento.

Por ocasião da Conferência Livre “Ciência Aberta no Brasil: desafios e oportunidades”, ocorrida no contexto da 5ª Conferência Nacional de Ciência, Tecnologia e Inovação (CNCTI), promovida pelo Ministério da Ciência, Tecnologia e Inovação (MCTI), em 22 de março de 2024, em que participei, o Doutor Paulo Artaxo (Sociedade Brasileira para o Progresso da Ciência - SBPC), evidenciava a ciência aberta como conjunto inclusivo de ações e práticas com objetivo de disponibilizar, tornar acessível e utilizável o conhecimento científico para benefício da sociedade. O cientista destaca a necessidade de integração de práticas acadêmicas de reprodutibilidade, compartilhamento, assim como a realização de pesquisas inclusivas, cooperativas e interinstitucionais, que promovam a popularização da ciência, universalidade do acesso ao conhecimento e reconhecimento da ciência aberta como motor de desenvolvimento econômico e social do país. Assim como, a importância de se discutir dados abertos no contexto da ciência aberta no Brasil, e conseqüentemente, a publicação de dados de pesquisa.

Nesse domínio, evidencia-se a ampliação das discussões sobre gestão de dados em nível mundial, justificando-se por diversas políticas e ações, desenvolvidas em diversos países, a exemplo do *Data Observatory (DO)*, maior observatório de dados de código aberto da Europa; o *Data Observatory*, colaboração público-privada-acadêmica visando maximizar os resultados para fornecimento de serviços baseados em dados para todo o ecossistema, centrados na coleta e limpeza de dados, armazenamento e integração de dados, visualização e Interpretação de Dados, análise, processamento e modelagem de dados e treinamento educação e suporte; O Observatório em Ciência, Tecnologia e Inovação em Saúde (Observatório CT&I em saúde) da Fundação Oswaldo Cruz (Fiocruz), que objetiva contribuir para a gestão da pesquisa e formulação de políticas institucionais em ciência, tecnologia, inovação e educação com uso de dados.

Diante de tal relevância, a presente obra intitulada “GESTÃO DE DADOS DE PESQUISA: Ampliando a geração de conhecimentos científicos no contexto da ciência aberta”, de autoria de Adriana Alves Rodrigues, Guilherme Ataíde Dias e Rafael Luís Galdini Raimundo, surge em um momento de efervescência das discussões que circundam o tema, inclusive, considerando os avanços advindos com as Tecnologias Digitais de Informação e Comunicação (TDICs). E, especialmente, levando-se em consideração dados de pesquisa que podem gerar novos conhecimentos e servir como base para variedade de novas pesquisas por diversos campos do saber. A referida obra, foi fruto do Projeto *DataPB*, de caráter transdisciplinar, cujo escopo corresponde à estruturação de um repositório para armazenar e compartilhar dados e materiais de pesquisa, em associação com outras ferramentas de ciência aberta e a colaboração entre pesquisadores, buscando fomentar o desenvolvimento de comunidade de pesquisa sustentável.

No segundo capítulo, seguindo a introdução, realiza-se uma abordagem sobre conjunto de dados gerados pela comunidade

científica, discorrendo sobre os conceitos que envolvem dados de pesquisa, destacando a variedade de conceitos e definições existentes sobre o assunto, deixando claro que a definição adotada considera a contribuição dos pesquisadores Sales e Sayão, como “todo e qualquer tipo de registro coletado, observado, gerado ou usado pela pesquisa científica, tratado e aceito como necessário para validar os resultados da pesquisa pela comunidade científica”. Apresentando ainda diretrizes da OCDE, obstáculos, benefícios do compartilhamento de dados, tipologia e o ciclo de vida dos dados, com ênfase no *Data Lifecycle Model*, proposto pelo *DataONE* e desenvolvido com apoio da *National Science Foundation* (NSF).

O terceiro capítulo, trata especificamente dos princípios usados na Gestão de Dados de Pesquisa. Apresenta os princípios FAIR relacionados ao acesso e reuso de dados de pesquisa e propõe avaliar a aderência a esses princípios, dos conjuntos de dados disponíveis, destacando a relevância dos recursos tecnológicos nos ambientes de pesquisa. Além disso, discorre sobre os Princípios CARE relacionados a governança de dados indígenas, visando o auxílio aos povos indígenas no tratamento, metadados, compartilhamento, uso e reuso dos dados abertos.

No quarto capítulo, aborda-se o plano de gestão de dados adotando estratégias que visam ações que possam garantir o gerenciamento e preservação adequados durante o processo de pesquisa. Apresenta como proposta do programa de gestão de dados a garantia de dados compreensíveis para outros pesquisadores ingressados no mesmo conjunto de dados, assim como, premissas e benefícios para o Plano de Gestão de Dados, observando-se a necessidade de que diretrizes e informações adicionais sejam disponibilizadas de forma clara e eficaz.

No capítulo cinco, discorre-se sobre a política de gestão de dados de pesquisa, como etapa fundamental para implementação de repositórios de dados, evidenciando a necessidade de formação de comitê gestor e a recomendação de uso de ferramentas de software específicos como o *DMPtool* ou o *DMP online* para elaboração do plano de gestão de dados. A proposta sugere que o modelo de política de gestão de dados envolva elementos como responsabilidades e autoridade, Coleta, Armazenamento de dados, Qualidade dos dados, Acessibilidade, Licenciamento, Integridade, Privacidade, Responsabilidade, Transparência, Aspectos Éticos e Período de embargo.

Por fim, expresso aos autores minha gratidão pela oportunidade, destacando que essa obra tem muito a contribuir para o campo da Ciência da informação e para o progresso da ciência no contexto de contributos para a disseminação da pesquisa científica, ampliação do acesso e, conseqüentemente, para o desenvolvimento socioeconômico nacional.

Rosilene Paiva Marinho de Sousa
Professora do Bacharelado em Direito
Universidade Federal do Oeste da Bahia - UFOB

AGRADECIMENTOS

À Fundação de Apoio à Pesquisa do Estado da Paraíba (FAPESQ/PB), Termo nº 1969/2022, referente a pesquisa de Pós-Doutorado desenvolvida no Programa de Pós-Graduação em Ciência da Informação na Universidade Federal da Paraíba (PPGCI/UEPB), assim como ao Edital FAPESQ/PB 09/2021, demanda universal.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico(CNPq), processo 311563/2018-0.

SUMÁRIO

1	INTRODUÇÃO.....	15
1.1	CONTEXTUALIZAÇÃO	15
2	DADOS DE PESQUISA.....	23
2.1	TIPOS DE DADOS	29
2.2	CICLO DE VIDA DE DADOS (CVD).....	32
3	PRINCÍPIOS USADOS NA GESTÃO DE DADOS DE PESQUISA.....	41
3.1	OS PRINCÍPIOS FAIR APLICADOS À GESTÃO DE DADOS	41
3.1.1	<i>FairDataBR: Avaliando a Aderência de Conjuntos de Dados aos Princípios FAIR.....</i>	45
3.2	OS PRINCÍPIOS CARE APLICADOS À GOVERNANÇA DE DADOS ÍNDIGENAS	47
4	PLANO DE GESTÃO DE DADOS	52

5	POLÍTICA DE GESTÃO DE DADOS DE PESQUISA	62
5.1	FERRAMENTAS ESPECIALIZADAS SUGERIDAS NO CONTEXTO DO REPOSITÓRIO DE DADOS	64
5.2	ESPECIFICAÇÕES PARA O REPOSITÓRIO DE DADOS	65
5.3	PRIVACIDADE E SEGURANÇA EM REPOSITÓRIOS DE DADOS	68
5.4	ATRIBUIÇÃO DE IDENTIFICADORES PERSISTENTES	69
6	CONSIDERAÇÕES FINAIS	73
	REFERÊNCIAS	75
	GLOSSÁRIO	83
	SOBRE OS AUTORES E AS AUTORAS.....	91



CAPÍTULO 1

1 INTRODUÇÃO

1.1 CONTEXTUALIZAÇÃO

Este trabalho aborda a questão da gestão dos dados de pesquisa no contexto da investigação científica, um tema cada vez mais premente entre os pesquisadores acadêmicos e profissionais. As atividades de gestão dos dados de pesquisa devem ser consideradas em praticamente todas as etapas e tipos de estudos, desde projetos de iniciação científica até estudos de mestrado, doutorado e pós-doutorado. Isso ocorre porque os dados produzidos durante o processo de investigação científica representam verdadeiras “pérolas”, com um potencial significativo para contribuir para a geração de novos conhecimentos quando submetidos a processos adequados de curadoria.

A ascensão dos dados de pesquisa está intrinsecamente ligada ao avanço e à rápida disseminação das Tecnologias Digitais de Informação e Comunicação (TDICs) a partir das últimas décadas do século XX. O uso generalizado de dispositivos digitais para gerar, coletar e processar dados de pesquisa tornou-se uma realidade onipresente, impulsionando o crescimento exponencial da disponibilidade de dados. É notável que não apenas os pesquisadores, mas também qualquer pessoa que faça uso de dispositivos tecnológicos, como celulares, cartões de crédito e *Smart TVs*, contribui para essa crescente abundância de dados. Esses dados, provenientes de uma variedade de fontes, representam uma fonte rica e primária para a geração de novos conhecimentos e são

essenciais para uma ampla gama de pesquisas. Vale ressaltar que a mesma tecnologia que facilita a geração de grandes volumes de dados é a mesma ferramenta que permite o seu processamento (Borgman, 2015).

Podemos compreender de maneira sintética a gestão de dados de pesquisa como uma sequência de atividades realizadas ao longo de todo um Ciclo de Vida dos Dados (CVD), com o objetivo de ampliar o uso, reuso e efetivo compartilhamento dos dados. Essa prática é fundamental, entre outras razões, porque permite a preservação dos dados a longo prazo. Destacamos que com frequência, os pesquisadores, ao término de um estudo, armazenam seus conjuntos de dados em dispositivos de armazenamento, como HDs ou SSDs, *pendrives*, *websites*, entre outros, o que dificulta o acesso a uma comunidade mais ampla. Esses conjuntos de dados estão em alto risco de serem perdidos, seja devido à obsolescência tecnológica ou a algum desastre imprevisto.

O desenvolvimento dessa obra deu-se no âmbito do Projeto *DataPB*¹, que se caracteriza por ser uma iniciativa de pesquisa transdisciplinar que busca estruturar um repositório para armazenar e compartilhar dados e materiais de pesquisa, em associação com outras ferramentas de ciência aberta e a colaboração entre pesquisadores. Especificamente, o projeto pretende fomentar o desenvolvimento de uma comunidade de pesquisa focada na co-construção da sustentabilidade a partir dos fundamentos teóricos e metodológicos providos por um bem consolidado arcabouço paradigmático constituído pela integração entre a teoria evolutiva e a teoria sócio-ecológica abordada sob uma visão de sistemas adaptativos complexos (*Datapb*, 2023).

Nesse contexto, a criação de um repositório de dados científicos articulado a projetos-piloto de reuso desses dados é o coração do

¹ Disponível em: <http://datapb.ccae.ufpb.br/>. Acesso em: 24 mar 2024

DataPB e está alinhada com uma rede de pesquisa, inovação e educação voltadas para a sustentabilidade no estado da Paraíba. Recentemente, essa iniciativa tem ganhado maior projeção em virtude de sua amplificação com a criação do Laboratório Misto Internacional sobre Sustentabilidade IDEAL (<http://ideal.ufpb.br>), o qual é liderado pela Universidade Federal da Paraíba em associação com o Instituto Francês de Pesquisa para o Desenvolvimento (IRD) e envolve um total de 11 instituições francesas e brasileiras. Nesse contexto, o repositório científico do *DataPB* (<https://dataverse.ideal.ufpb.br>) está servindo como alicerce fundamental de uma ciber-infraestrutura com potencial para uso em diferentes estados do Nordeste brasileiro.

O objetivo do Repositório de Dados de Pesquisa *DataPB* é oferecer à comunidade científica e acadêmica das instituições de ensino e pesquisa localizadas na Paraíba uma infraestrutura virtual para depósito e compartilhamento de dados de pesquisa. A prioridade é para os dados gerados por pesquisadores atuantes na área de biodiversidade e de várias outras áreas que geram informações quantitativas estratégicas para a pesquisa sobre sustentabilidade, tais como economia, política e ciências sociais. Entretanto, estão excluídos desta iniciativa os dados de pesquisa que apresentem restrições de confidencialidade, conforme justificado de acordo com a legislação nacional em vigor (Fundação Oswaldo Cruz, 2020).

O avanço da ciência contemporânea está intrinsecamente relacionado ao progresso e à aplicação das TDICs, os quais mantêm uma estreita associação com os processos de construção do conhecimento, os quais são intensificados pelo uso dessas tecnologias. Essa conjunção de fatores proporciona oportunidades sem precedentes para impulsionar o empreendimento científico, especialmente diante da urgente necessidade de desenvolver inovações que possam enfrentar os desafios prementes do século XXI (Hey *et al.*, 2009).

A crescente produção de volumes consideráveis de dados e sua intensiva utilização, impulsionados pela rápida evolução tecnológica, destaca a necessidade urgente de desenvolver abordagens eficazes para gerenciar os processos subjacentes ao uso, reúso e disseminação de dados. Exemplos notáveis dessa revolução em curso, resultante do amplo uso de dados, incluem diversas estratégias de aplicação da Inteligência Artificial (IA). Entretanto, o potencial dessas abordagens só poderá ser plenamente realizado pela disponibilidade de grandes conjuntos de dados organizados de forma a permitir sua exploração completa. As transformações observadas no âmbito das pesquisas científicas são diretamente influenciadas pelo acesso aos conjuntos de dados *online* e pela possibilidade de colaboração entre pesquisadores ao redor do mundo, viabilizada pela infraestrutura disponibilizada pela Internet. Nesse contexto, o aumento significativo no uso de vastos e diversificados volumes de dados, identificados como *big data*, representa um marco importante na evolução da ciência. Isso se caracteriza pelo emprego de conjuntos de dados em larga escala e pela utilização de recursos tecnológicos para sua gestão eficiente.

De acordo com os autores Mayer-Schönberger; Cukier, 2013, p. 8), o *big data* se constitui em “uma nova fonte de valor econômico e informação” como também “marca um grande passo na humanidade por quantificar e compreender o mundo” (Mayer-Schönberger; Cukier, 2013, p.11). No campo científico, essa abordagem está relacionada com o conceito de *e-Science*, cujo termo foi desenvolvido por John Taylor no Reino Unido em 1999, quando era diretor geral do *Research Councils UK* (Jankowski, 2007). Ele se refere ao conjunto de instrumentos e aparatos tecnológicos adequados para auxiliar a ciência contemporânea diante da produção dos grandes volumes de dados produzidos diariamente, com características colaborativas e multidisciplinares.

A *e-Science* é considerada como o quarto paradigma da ciência (Gray, 2007). Para Lippincott (2010, p.63), a terminologia é comumente

explorada para “descrever a emergência da ciência orientada a dados (data-driven *science*)”. Ao analisar as práticas colaborativas da *e-Science*, Voss *et al.* (2010, p. 4165, tradução nossa)² adotam o conceito de aprendizagem social como forma de refletir sobre o contexto contemporâneo. Para os autores, a aprendizagem social se refere à compreensão de como as competências tecnológicas são apropriadas e incorporadas na sociedade, cujo movimento está diretamente vinculado aos processos de inovação e da variedade de atores envolvidos.

A perspectiva da aprendizagem social sugere uma visão do desenvolvimento e uso de tecnologias que considera todo um espectro de atividades que vão desde a interação imediata com tecnologias em ambientes locais (aprender fazendo) até a troca de experiências em amplas redes de design e uso de tecnologia (aprendendo interagindo). (Voss *et al.*, 2010, p. 4165, tradução nossa)

Contudo, os autores fazem uma ressalva de que esta integração e apropriação tecnológica pelos atores envolvidos deve ser domesticada para seu uso, e que essa domesticação não ocorre em um processo marcado pela linearidade e está longe de ter um ponto final, apenas encerramentos momentâneos, uma vez que esse movimento ocorre no contexto da inovação, mas que podem ser reabertos com vistas a novas desenvolvimentos.

Nessa trajetória, as capacidades das TDICs no contexto da ciência aberta têm incentivado mudanças estruturais nas práticas de pesquisa científica. Surge, então, a necessidade de uma gestão dinâmica dos dados de pesquisa, abordando todas as etapas, da coleta ao compartilhamento desses dados. A disponibilidade de serviços de

² No original: “The social learning perspective suggests a view of the development and use of technologies that considers a whole spectrum of activities ranging from the immediate interaction with technologies in local settings (learning by doing) to the exchange of experiences in wide networks of technology design and use (learning by interacting)”.

gestão de dados emergiu em resposta às transformações advindas do quarto paradigma da ciência, que demandou uma ciberinfraestrutura capaz de suportar o processamento dos dados, gerenciar ciclos de vida e estabelecer políticas de gestão de dados de pesquisa (Chiware & Mathe, 2015; Anjos, Dias & Rodrigues, 2017).

A implementação de uma política de gestão de dados de pesquisa refere-se a uma coleção de regras institucionais que determinam as maneiras de tratar, curar, armazenar e reutilizar os dados, assim como de depositar e gerenciar os dados originários de pesquisas. Conforme destaca Erway (2013, p.7, tradução nossa)³,

uma política de toda a universidade deve abordar as melhores práticas no gerenciamento de dados de pesquisa e torná-las acessíveis ao público (quando viável), contribuindo para a pesquisa de alta qualidade, integridade acadêmica e administração responsável

O desenvolvimento de critérios para compartilhar, disponibilizar, usar e reutilizar dados de pesquisa tornou-se uma necessidade primordial nas instituições de pesquisa, pois a sistematização e manutenção de uma infraestrutura para a gestão de dados envolvem questões potencialmente controversas.

Uma política de gestão de dados de pesquisa requer a elaboração de um documento formal que funcione como um guia para a gestão eficaz dos dados. Esse documento deve ser composto por princípios normativos que orientem a sua implementação. É fundamental que as premissas normativas contidas no documento sejam adaptadas à realidade de cada instituição de ensino e pesquisa, considerando as esferas culturais, tecnológicas e sociais das comunidades acadêmicas

³ No original: “A university-wide policy should address best practices in managing research data and making them accessible to the public (where feasible), contributing to high-quality research, academic integrity, and responsible administration”.

e científicas. Além disso, é essencial promover discussões para compreender o funcionamento da cadeia produtiva da infraestrutura dos dados de pesquisa.



CAPÍTULO 2

2 DADOS DE PESQUISA

Desde as últimas décadas do século XX, os procedimentos inerentes à condução de investigações científicas foram caracterizados por uma crescente adoção das TDICs em todas as áreas do conhecimento. Este fenômeno implica que um número substancial de pesquisadores atualmente produz e utiliza conjuntos de dados volumosos como pilares fundamentais de suas pesquisas. Diante desse crescimento contínuo, surge a necessidade urgente de desenvolver novas metodologias e capacitar os pesquisadores para lidar eficazmente com esses conjuntos de dados. Os dados gerados pela comunidade científica abrangem uma variedade de formatos, englobando desde estatísticas e resultados experimentais até gravações e transcrições de entrevistas (Borgman, 2012).

Antes de prosseguirmos, é importante destacar algumas definições para o conceito de dados. A definição do que constitui um dado não é universalmente estabelecida; na literatura, é possível encontrar diversas interpretações. Nesse sentido, trazemos o pensamento de Santos e Sant’Ana (2013, p. 201):

Muitas foram as tentativas de definir o conceito de dado, e uma das dificuldades inerentes à definição reside na condição do termo e do próprio dado serem utilizados de maneiras distintas por praticamente todas as áreas da ciência.

No contexto deste trabalho, apresentamos duas definições de dados elaboradas por pesquisadores brasileiros que consideramos

adequadas dentro do escopo dos temas abordados no domínio da gestão de dados de pesquisa, embora existam várias outras disponíveis.

No campo da Ciência da Informação, destacamos a contribuição de Santos e Sant’Ana (2002, p.3) ao considerarem “*dado* como um elemento básico, formado por signo ou conjunto finito de signos que não contém, intrinsecamente, um componente semântico, mas somente elementos sintáticos.”. Uma definição simples e de caráter operacional e computacional é apresentada pelo professor aposentado da Universidade de São Paulo (USP) e pioneiro da Ciência da Computação Brasileira, Valdemar W. Setzer, o referido professor explica: “Defino dado como uma seqüência de símbolos quantificados ou quantificáveis” (Setzer, 2015, *online*).

Embora definir dados possa apresentar desafios, Borgman (2015, *kindle location*, 618, tradução nossa)⁴ afirma que a definição de dados se encontra em uma fronteira pouco clara sobre o que é ou não é considerado dado. Para a pesquisadora,

dados são formas de informação, um amplo conceito que está cada vez mais difícil definir. Problemas epistemológicos e ontológicos são abundantes, resultando em muitos livros dedicados à explicação de informação e conhecimento.

A autora destaca que, apesar da falta de clareza semântica, o conceito de dados tem ganhado importância nas pesquisas de humanidades digitais e, que, ao mesmo tempo, os pesquisadores cada vez mais vem produzindo um “dilúvio” (*deluge*) de dados com novas formas metodológicas e de instrumentalização para as pesquisas científicas (Borgman, 2012).

4 No original: “Data is a form of information, a broad concept that is increasingly difficult to define. Epistemological and ontological problems abound, resulting in many books devoted to the explanation of information and knowledge.”

No que diz respeito ao conceito de dado de pesquisa, é possível identificar uma ampla variedade de definições sobre o assunto. Para a Organização para Cooperação e Desenvolvimento Econômico (OCDE, 2007, p.13), estes são

registros factuais (escores numéricos, registros textuais, imagens e sons) utilizados como fontes primárias para pesquisas científicas, e que são comumente aceitos na comunidade científica como necessário para validar resultados de pesquisa

Dados de pesquisa também podem ser compreendidos como todo material que resulta do processo de coleta ou é proveniente de dados primários, quantitativos ou qualitativos, bem como dados que sejam originários de “fontes pré-existentes e que são passíveis de análises no escopo de uma pesquisa determinada” (Grant, 2017, *online*).

Ao longo deste texto, adotaremos como definição de referência a contribuição dos pesquisadores Sales e Sayão (2013, p. 3), os quais conceituam dados de pesquisa como:

”

Todo e qualquer tipo de registro coletado, observado, gerado ou usado pela pesquisa científica, tratado e aceito como necessário para validar os resultados da pesquisa pela comunidade científica.

[Sales e Sayão 2013, p. 3]

“

O paradigma emergente, no qual a utilização de conjuntos de dados se configura como uma prerrogativa fundamental para a geração de novos conhecimentos, requer que os pesquisadores realizem uma gestão adequada dos conjuntos de dados de pesquisa produzidos. Isso se revela importante para assegurar o acesso, preservação e viabilidade de uso, reutilização e compartilhamento desses conjuntos. Assim, estabelece-se uma relação simbiótica entre pesquisadores, instituições e agências de fomento.

Os dados de pesquisa têm desempenhado um papel fundamental na ciência moderna, destacando-se pelo compartilhamento e impulsionando o desenvolvimento de novos métodos e ferramentas inovadoras capazes de lidar com grandes volumes de dados. Nesse contexto, ao examinar o ambiente de pesquisa permeado por tecnologias computacionais, Gray (2007) propõe uma estrutura informacional na forma de pirâmide (Figura 1), na qual os dados se encontram na base e a literatura no topo.

Figura 1 - Pirâmide Informacional dos dados de pesquisa



Fonte: Elaboração dos autores com base em Gray (2007)

Como uma das características da ciência aberta, é importante observar que os dados brutos ocupam a base da pirâmide no universo das pesquisas, e seu compartilhamento é uma das etapas primordiais nas pesquisas científicas, a partir do chamado quarto paradigma científico. Os dados se tornaram essenciais no ecossistema da *e-Science* e são fundamentais para a criação de novas fontes de pesquisa, conhecimento, investigação, métodos e inovação tecnológica.

Para Borgman (2015), o compartilhamento de dados é uma questão urgente. Embora não seja um procedimento novo, os dados estão adquirindo características de ubiquidade e pervasividade na sociedade contemporânea. A compreensão do fenômeno depende da ambiência na qual os dados estão inseridos e da perspectiva do pesquisador.

Os dados não são objetos puros ou naturais com essência própria. Eles existem em um contexto, adquirindo significado a partir desse contexto e formando a perspectiva de quem os observa (Borgman, 2015, p.18, tradução nossa⁵).

Além disso, o compartilhamento de dados pode assumir diferentes formas, como troca privada mediante solicitação, coleta de dados, postagem de conjuntos de dados em diversos repositórios e fornecimento de material suplementar.

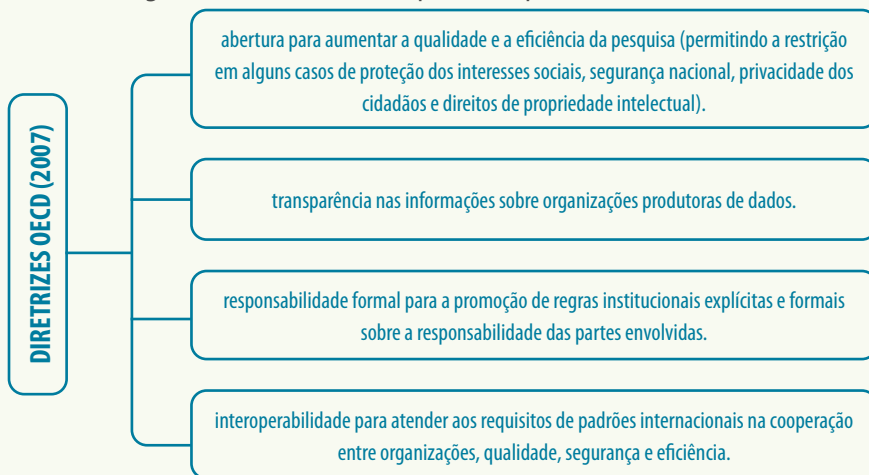
A prática do compartilhamento de dados de pesquisa no cenário da *e-Science* tem o objetivo de acelerar a produção de novas descobertas científicas, otimizar o uso de recursos tecnológicos e financeiros, tornar o processo mais transparente e ampliar as possibilidades de verificação e

⁵ No original: “Data are not pure or natural objects with an essence of their own. They exist in a context, taking on meaning from that context and form the perspective of the beholder”.

reuso das pesquisas científicas, além de oferecer um retorno à sociedade no caso de pesquisas financiadas com recursos públicos.

Diversas instituições internacionais e agências de fomento têm proposto a elaboração de Planos de Gestão de Dados (PGDs), objetivando maximizar o compartilhamento desses insumos, como, por exemplo, o Comitê de Política Científica e Tecnológica da OCDE, que estabeleceu os “Princípios e Diretrizes para Acesso a Dados de Pesquisa com Financiamento Público” em 2007. Essas diretrizes (Figura 2) reforçam a importância da divulgação dos dados científicos em países em desenvolvimento, contribuindo para a formação de novos pesquisadores em nível global. O documento enumera ainda uma série de preceitos que fundamentam o acesso aberto aos dados.

Figura 2 - Diretrizes da OCDE para o compartilhamento de dados



Fonte: Elaboração dos autores com base na OCDE (2007)

As práticas advindas do compartilhamento de dados integram aspectos nas esferas sociais e tecnológicas que se configuram como questões complexas na ciência contemporânea. Para Veiga, Silva e Borges (2021) o compartilhamento de dados se torna um fator

primordial no provimento e acesso aos dados de pesquisa que apresentam processos desafiantes para sua prática. Os autores elencam os aspectos desafiantes no que se refere a essa prática. Além disso, é preciso levar em conta as causas individuais e institucionais que podem dificultar o compartilhamento de dados. Os autores elencam alguns fatores que influenciam compartilhamento dos dados de pesquisa. Esses fatores podem ser classificados em custos, benefícios e fatores contextuais, tanto individuais quanto institucionais (Veiga, Silva e Borges, 2021). A interação entre esses fatores determina, em grande medida, a disposição dos pesquisadores em disponibilizar seus dados para a comunidade científica.

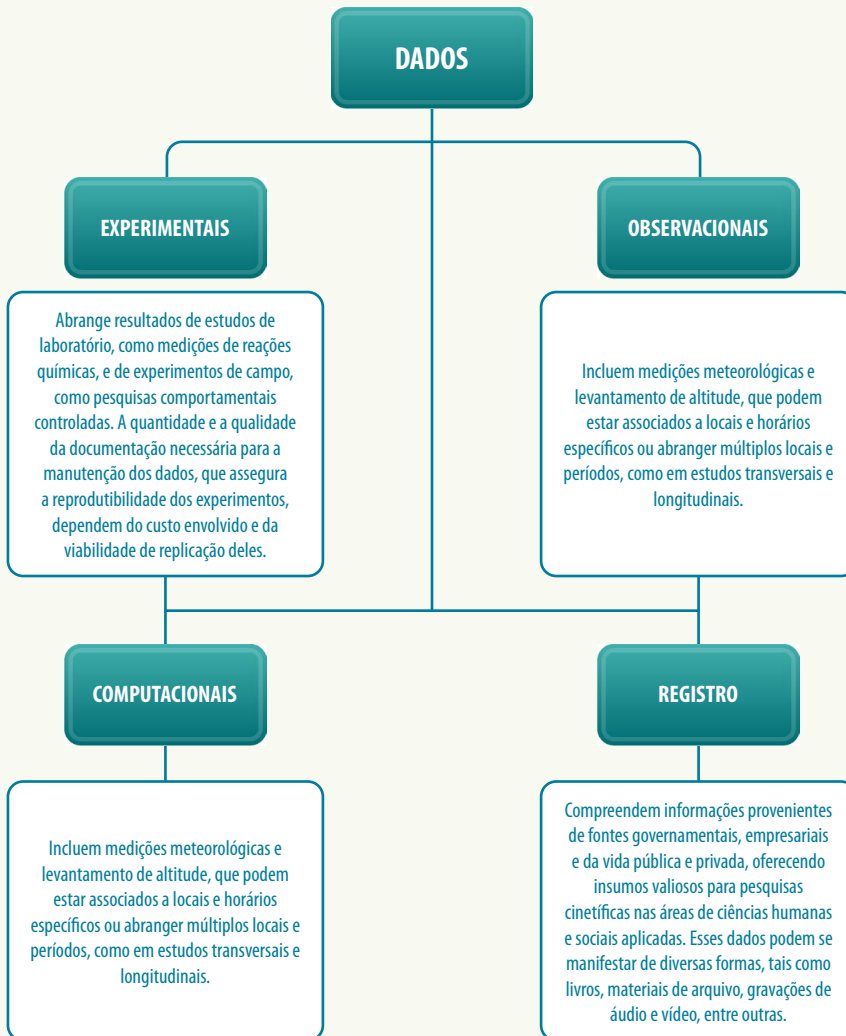
Neste contexto, além de estabelecer os aspectos individuais e institucionais, os autores também apontam alguns benefícios para o compartilhamento de dados de pesquisa, a saber: Acessibilidade, Visibilidade, Credibilidade, Recompensa Acadêmica, Reconhecimento Profissional e por fim, Altruísmo. Ainda nesta perspectiva, é preciso que sejam implementadas diretrizes, estratégias educativas e comunicativas para o funcionamento produtivo entre pesquisadores e plataformas de dados como instrumento de mediação acadêmica e compartilhamento dos saberes científicos. Deste modo, a inserção de uma gestão de dados de pesquisa pode garantir o acesso aos dados, compartilhamento, uso e recursos com vistas aos avanços científicos e de pesquisas (Veiga, Silva e Borges, 2021)

2.1 TIPOS DE DADOS

No que tange à tipologia, os dados estavam tradicionalmente estruturados em três categorias: dados experimentais, dados computacionais e dados observacionais. Posteriormente foi incluído

uma quarta categoria, dados do tipo registro (Borgman, 2012). A Figura 3 ilustra essa tipologia.

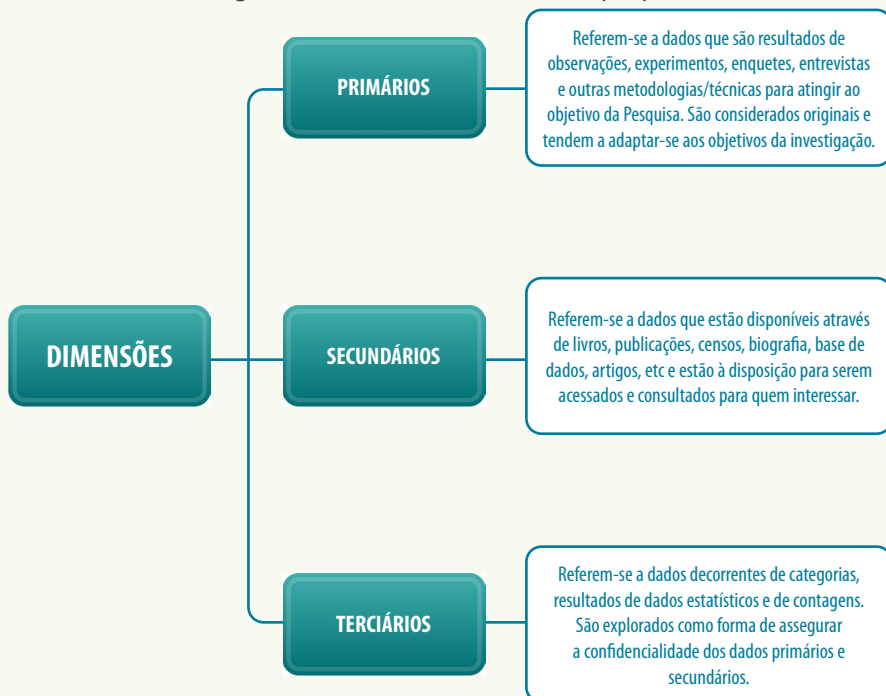
Figura 3 - Tipologia dos dados



Fonte: Elaboração dos autores com base em Borgman (2015)

Conforme a tipologia dos dados apresentada na Figura 3, é preciso levar em conta a singularidade de cada tipo de dado, ou seja, cada dado tem uma estrutura própria que o individualiza. Tal cenário da diversidade dos dados vai requerer políticas institucionais bem delineadas que abarquem essas características e demais dinâmicas associadas em todo o processo de construção das políticas. A diferenciação dos dados pela *National Science Board* (NSB) conduz à demanda de considerar tipos de políticas diferentes, bem como sistemas e serviços que deem conta da heterogeneidade dos dados, e uma adequação da ciberinfraestrutura. Em conformidade com Silva (2019), os dados apresentam outras características (Figura 4) levando em conta a sua natureza, a saber:

Figura 4 - Características dos dados de pesquisa



Fonte: Elaboração dos autores com base em Silva (2019)

Conforme delineado por Sales e Sayão (2013), a compreensão da relevância dos dados nas investigações científicas, abrangendo aspectos relacionados ao seu tratamento, preservação e gestão, já se manifesta em práticas acadêmicas observadas em instituições de fomento e estabelecimentos de ensino e pesquisa, entre outros contextos.

2.2 CICLO DE VIDA DE DADOS (CVD)

De modo geral, a Gestão de Dados de Pesquisa pode ser compreendida como um arsenal de procedimentos que requer ferramentas tecnológicas e infraestruturas adequadas, englobando serviços interligados nas fases do ciclo de vida dos dados (Dudziak, 2016). Esse procedimentos podem ser disponibilizados, em sua maioria, pelas instituições acadêmicas, agências de fomento ou entidades vinculadas ao desenvolvimento das pesquisas, que começaram a modificar a forma de administrar os dados com o estabelecimento de políticas mandatórias visando uma melhor gestão dos dados de pesquisa (Chiwere; Mathe, 2015; Anjos; Dias; Rodrigues, 2017).

Segundo Whyte e Tedds (2011, p.1, tradução nossa), gestão de dados está relacionado “à organização de dados, desde sua entrada no ciclo de pesquisa até a disseminação e arquivamento de resultados valiosos”⁶. Esta definição se refere a um agrupamento de práticas e processos diferenciados que estão envolvidos nos seguintes aspectos: design, criação dos dados, segurança, armazenamento, recuperação, preservação, reutilização e compartilhamento dos dados considerando as questões de capacidades tecnológicas, questões éticas, legislação e arcabouço de gerência.

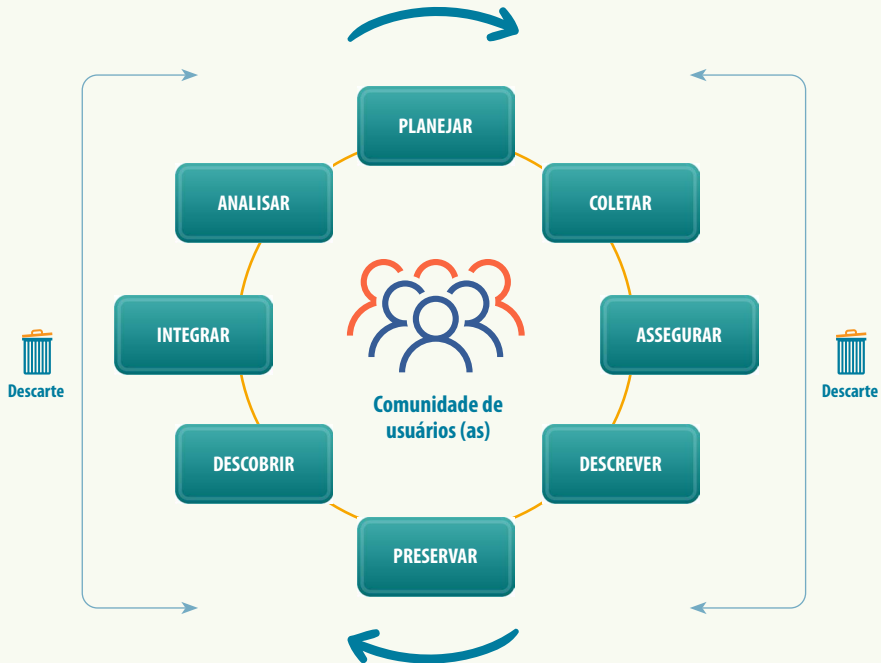
⁶ No original: “to the organization of data, from its entry into the research cycle to dissemination and archiving to valuable outputs”.

Simultaneamente à geração de um enriquecimento de dados de pesquisa, os quais assumiram um papel central na ciência moderna, surgiram desafios no que concerne às práticas de gestão desses dados. As universidades estão agora empenhadas em identificar as melhores abordagens e ferramentas necessárias para a administração dos dados de pesquisa, incluindo considerações sobre curadoria, questões técnicas e armazenamento, entre outros aspectos. Uma gestão eficaz deve incorporar premissas essenciais com o propósito de facilitar um compartilhamento produtivo entre a comunidade científica e outros grupos interessados que possam se beneficiar desses dados, especialmente aqueles provenientes de observações únicas.

A gestão de dados tornou-se essencial no contexto atual das instituições de pesquisa, visto que grandes volumes de dados de pesquisa digitais são gerados constantemente em diversas formas e a um ritmo acelerado. Essa realidade impõe os desafios dos 3Vs do *Big Data*: “Volume”, “Variedade” e “Velocidade”. Para enfrentar esses desafios, é essencial adotar uma abordagem estruturada, como o emprego do Ciclo de Vida dos Dados. Um modelo amplamente citado na literatura é o *Data Lifecycle Model*, proposto pelo *DataONE* e desenvolvido no contexto da *National Science Foundation (NSF)*⁷, conforme representado na Figura 5.

⁷ A Fundação Nacional da Ciência (NSF) é uma agência independente do Governo Americano que promove a pesquisa e educação fundamental em todos os campos da ciência e engenharia.

Figura 5 - Modelo de Ciclo de Vida dos Dados do DataONE



Fonte: Elaboração dos autores com base em Dataone (2023, *online*, tradução nossa)

Este modelo demonstra uma síntese geral e a interrelação das fases do *DataONE Data Lifecycle Model*. O modelo inclui oito etapas que podem fazer parte de qualquer projeto que inclua o uso de dados, mas pode ser adaptado a vários domínios ou comunidades a partir das etapas, quais sejam:

- **Planejar:** inclui decisões sobre como os dados serão coletados, gerenciados, descritos e tornado acessível durante toda a sua vida (Dataone, 2023, *online*, tradução nossa);
- **Coletar:** refere-se às atividades de coleta de dados usando observações, instrumentação e outros métodos (Dataone, 2023, *online*, tradução nossa);

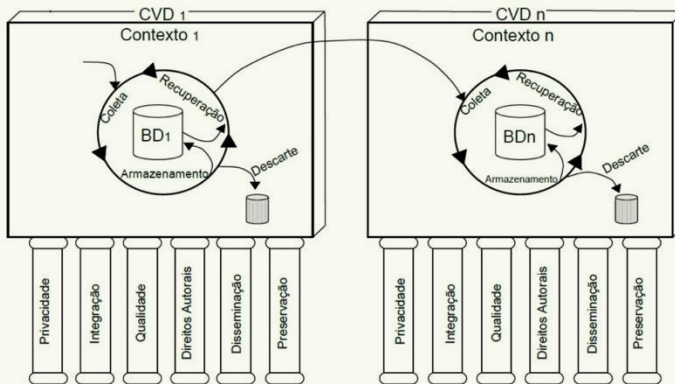
- **Garantir:** refere-se a garantia do status de qualificação dos dados por meio de verificações e inspeções que são aceitas dentro de uma determinada pesquisa comunidade (Dataone, 2023, *online*, tradução nossa);
- **Descrever:** inclui o registro de todas as informações técnicas, contextuais, administrativas e informações científicas sobre dados (metadados) usando os padrões de metadados apropriados (Dataone, 2023, *online*, tradução nossa);
- **Preservar:** inclui preparar e o envio dos dados para um repositório apropriado de longo prazo ou Centro de dados (Dataone, 2023, *online*, tradução nossa);
- **Descobrir:** refere-se a atividades que visam fornecer ferramentas e metadados abordagens para encontrar e recuperar dados e usar essas ferramentas para localizar e usar dados em a pesquisa de alguém (Dataone, 2023, *online*, tradução nossa);
- **Integrar:** refere-se a acessar várias fontes de dados e combinar dados para formar conjuntos de dados que podem ser analisados dentro do conjunto específico de questões ou hipóteses de pesquisa (Dataone, 2023, *online*, tradução nossa);
- **Analisar:** envolve o uso de várias ferramentas, métodos e técnicas para analisar dados. (Dataone, 2023, *online*, tradução nossa)⁸.

A análise da relevância do Ciclo de Vida dos Dados na gestão de dados de pesquisa na Ciência da Informação (CI) por Sant'ana (2016) enfatiza como a CI pode contribuir para o desenvolvimento de uma estrutura orgânica no uso de dados. O autor propõe seis fases

⁸ No original: <https://towson.libguides.com/dataresearch/datamangement/dldataone>. Acesso em: 21 mar. 2024.

envolvendo os dados: Privacidade, Integração, Qualidade, Direitos autorais, Disseminação e Preservação. Essas etapas estão inseridas no contexto de Coleta, Recuperação, Armazenamento e Descarte de Dados. A Figura 6 ilustra tal cenário de forma esquemática.

Figura 6 - Modelo de Ciclo de Vida dos Dados na Ciência da Informação



Fonte: Sant'ana (2016, p.123).

Este modelo, conforme explicado pelo autor, tem como objetivo maior possibilitar caminhos para a constituição de uma estrutura para a gestão de dados que leve em conta aspectos como:

esforços, estudos e ações com vistas à obtenção, manutenção e uso dos dados, tornando possível aproximar elementos semelhantes e distribuir teorias e metodologias em função de seu escopo seja por fase ou por fatores identificados no modelo de ciclo de vida dos dados (Sant'ana, 2016, p. 138).

Os dados de pesquisa também estão vinculados aos aspectos administrativos das universidades vinculadas. Sales e Sayão (2019, p.8-9) sugerem a criação de um comitê gestor com a capacidade de gerenciar os dados de pesquisa em três instâncias: 1) De modo Administrativo e

de suporte; 2) Pesquisa e gerenciamento; e 3) Aprendizagem e Ensino. Os autores acreditam que desta forma haverá a implantação de uma visão organizativa da gestão de dados, o que pode facilitar a inserção das políticas. Para isso, os membros dos comitês incluem esses papéis e funções, a saber:

- **Pesquisador** – principal agente envolvido com a pesquisa e com os dados; como autor/criador/coletor/avaliador dos dados/avaliador deve garantir que os metadados disciplinares, o registro dos dados (proveniência), a documentação, o contexto e a qualidade estejam em conformidade com os padrões estabelecidos pela comunidade ou pela instituição (Sales; Sayão, 2019);
- **Bibliotecário de dados** – profissional de biblioteconomia especializado em gestão de dados, responsável por catalogar, indexar, organizar e apoiar a publicação de conjuntos de dados; assessora o planejamento e a operacionalização dos repositórios e dos serviços de gestão de dados; apoia a curadoria através da construção de instrumentos de representação e padronização; idealmente o bibliotecário de dados conhece os fluxos de pesquisa de sua instituição; promove cursos, divulgação e material didático e assessora os pesquisadores na elaboração do plano de gestão de dados (Sales; Sayão, 2019);
- **Data steward**⁹ – profissional responsável pela gestão sustentável de dados de pesquisa a longo prazo, abrangendo

⁹ “Marín-Arraiza & Vidotti (2019) trabalharam o conceito de *data steward*, traduzindo, o termo como administrador de dados. De acordo com a visão das autoras, esse profissional seria o grande gestor que está na alta direção e promove a gestão de dados através da formulação de políticas. No entanto, os autores da presente pesquisa entendem que o conceito de *data steward*, conforme Mons (2018) que destaca que a principal habilidade desse profissional “é “ajudar os especialistas do domínio a discernir padrões significativos, correlações verdadeiras e, mais importante, a cavar as explicações mecanicistas e as relações causais que levam ao conhecimento acionável” (Mons, 2018, p.11, tradução nossa) e define *data steward* como “tratar os dados e os objetos de pesquisa associados com o máximo cuidado, com o objetivo de torná-los reutilizáveis para

todas as fases, desde o planejamento do estudo, coleta, análise, armazenamento, até o compartilhamento dos dados. Suas funções envolvem todas as atividades necessárias para assegurar que os dados digitais de pesquisa sejam localizáveis, acessíveis, interoperáveis e reutilizáveis (princípios FAIR), o que inclui o gerenciamento de dados, seu arquivamento e o reuso por terceiros (Sales ; Sayão, 2019);

- **Arquivista de dados** – profissional da área de arquivologia, encarregado do arquivamento e da preservação de longo prazo dos dados, assegurando a sua integridade, autenticidade e confiabilidade. Apoia o planejamento de sistemas de arquivamento confiáveis (Sales; Sayão, 2019);
- **Cientista de dados** – profissional especializado em Ciência da Computação e/ou em uma área disciplinar específica, que contribui para o desenvolvimento de tecnologias de análise, manipulação, visualização, modelagem, algoritmização de dados. Aplica também metodologias avançadas, como inteligência artificial e aprendizagem de máquina. (Sales; Sayão, 2019);
- **Gerente de dados** – profissional da área de tecnologia da informação responsável pela manutenção e implementação de bases de dados, repositórios, sistemas de armazenamento. Suas funções englobam o suporte à segurança dos dados, a realização de *backups* e a verificação da integridade dos dados (Sales; Sayão, 2019);
- **Curador de dados** – Pesquisador ou profissional da informação com conhecimento especializado na área disciplinar, que agrega valor aos dados por meio da criação de documentação, elaboração de metadados, atribuição de

descoberta, desde que sejam válidos” (Mons, 2018, p. 24, tradução nossa)” (Sales e Sayão, 2019, p. 22).



identificadores, contextualização, integração, reformatação e *mashup*, entre outras atividades. Este profissional também promove o compartilhamento e o reuso dos dados e contribui para a avaliação necessária à sua preservação e ao desenvolvimento de serviços relacionados (Sales; Sayão, 2019);

- **Gestor** – administrador de Ciência e Tecnologia (C&T) que compreende a importância estratégica dos dados nos níveis institucional, nacional e internacional. Esse profissional contribui para a formulação de políticas de dados, negocia recursos junto a agências de fomento, implementa e-infraestruturas e adquire ferramentas, equipamentos, produtos de *software* e coleções de dados essenciais para a gestão eficiente dos dados de pesquisa. (Sales; Sayão, 2019).

Os papéis e funções dos membros que compõem o comitê gestor supramencionado são sugeridas e podem ser adaptadas para atender de forma mais eficaz às necessidades específicas de cada caso.

CAPÍTULO 3

3 PRINCÍPIOS USADOS NA GESTÃO DE DADOS DE PESQUISA

3.1 OS PRINCÍPIOS FAIR APLICADOS À GESTÃO DE DADOS

Os princípios FAIR (*Findable, Accessible, Interoperable and Reusable*) foram concebidos durante a Conferência Internacional “Jointly designing a data FAIRPORT” em Leiden, Holanda, em 2014. O objetivo principal desses princípios era abordar os desafios relacionados ao acesso e reuso de dados de pesquisa. Neste evento, acadêmicos, pesquisadores e outros interessados se reuniram para desenvolver um conjunto de diretrizes e atividades destinadas a facilitar a descoberta, acesso, integração e reutilização apropriada dos dados de pesquisa pela comunidade acadêmica. Como definição operacional para o acrônimo FAIR, destacamos a proposta de Rodriguez-Iglesias *et al* (2016, p.1, tradução nossa)¹⁰.

¹⁰ No original: “: FAIR is an acronym of Findable, Accessible, Interoperable, and Reusable. Briefly, the FAIR Principles suggest that every data element should have a globally-unique identifier, and that this identifier should be associated with contextual, searchable metadata (“Findable”); these identifiers should all resolve to data or metadata using an open, standard protocol (“Accessible”); the data and metadata should use a formal, broadly applicable representation language, and utilize open and widely-accepted domain-relevant vocabularies and ontologies (“Interoperable”); and finally, the data should be richly described with an abundance of cross-references, and with a clearly-defined mechanism for accessing provenance and license information (“Reusable”).

”

FAIR é um acrônimo de Encontrável, Acessível, Interoperável e Reutilizável. Resumidamente, os Princípios FAIR sugerem que cada elemento de dados deve ter um identificador único global, que deve ser associado a metadados contextuais e pesquisáveis ('Encontrável'). Esses identificadores devem resolver dados ou metadados usando um protocolo aberto, padrão ('Acessível'); os dados e os metadados devem usar uma linguagem de representação formal, amplamente aplicável, e vocabulários e ontologias abertas e amplamente aceitas para o domínio relevante ('Interoperável'); e, finalmente, os dados devem ser ricamente descritos com referências cruzadas abundantes e com um mecanismo claramente definido para acessar informações de proveniência e licença ('Reutilizável').

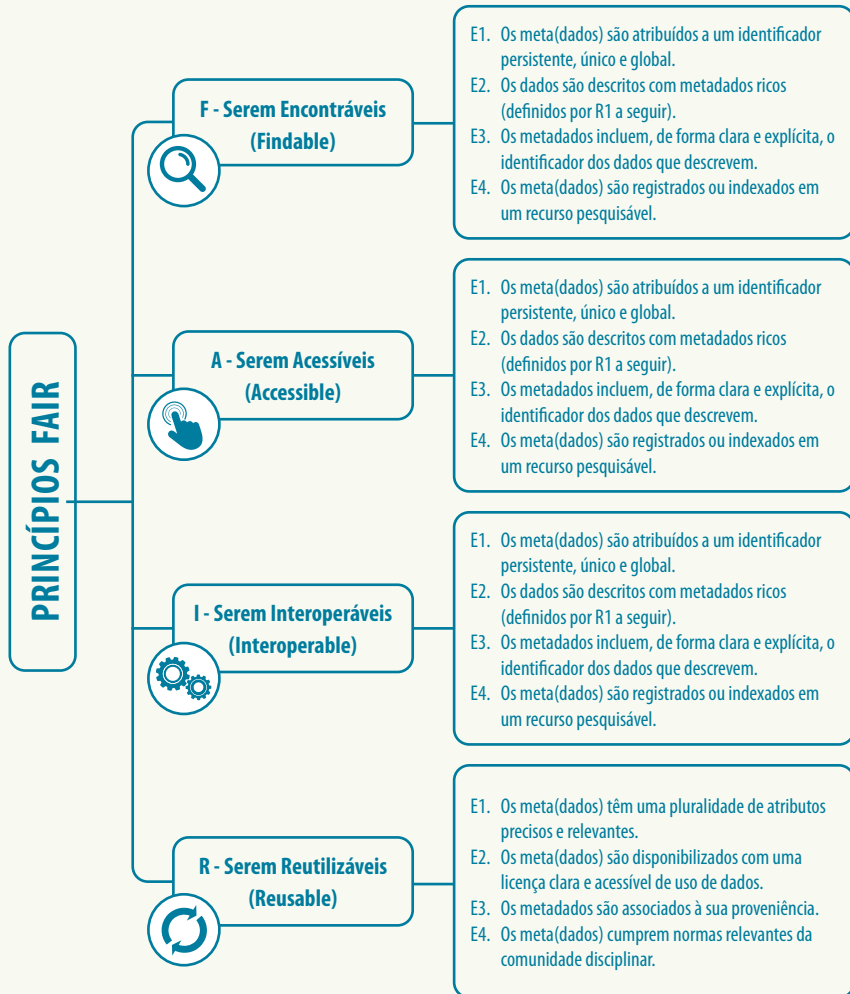
[RODRIGUEZ-IGLESIAS et al, 2016, p.1]

“

Os princípios FAIR ressaltam a importância das ferramentas tecnológicas nos ambientes de pesquisa ricos em dados, especialmente a prática de gerenciamento de processos por meio da inteligência artificial, em uma relação simbiótica que auxilia todo o processo de pesquisa, potencialmente reduzindo erros no manejo dos dados.

Um aspecto fundamental no contexto dos princípios FAIR é sua capacidade de serem ativados por máquinas (*machine actionability*), já que isso pode contribuir de maneira mais eficaz através do processamento por entidades não humanas, dada a enorme quantidade de conjuntos de dados atualmente disponíveis. A Figura 7 mostra os princípios FAIR e suas características.

Figura 7 - Os princípios orientadores FAIR



Fonte: Elaboração dos autores com base em Wilkinson *et al* (2016, *online*)

O princípio FAIR Encontrável (*Findable*) aborda a capacidade de aumentar a encontrabilidade de conjuntos de dados, como sugere o próprio nome, sendo essa uma premissa fundamental para a implementação dos demais princípios. Nesse contexto, os dados devem

ser descritos por meio de metadados detalhados para facilitar a busca e o acesso do pesquisador a esses dados (Dias; Anjos; Rodrigues, 2019). O princípio Acessível (*Accessible*) ressalta a capacidade de disponibilizar e permitir acesso aos dados quando são encontrados, considerando os aspectos relacionados à autenticação e autorização.

O princípio FAIR Interoperável (*Interoperable*) está associado à possibilidade de integrar os dados a outros conjuntos de dados, levando em consideração os metadados e estabelecendo conexões com outras instâncias relacionadas ao Ciclo de Vida dos Dados. Para isso, é fundamental a existência de ferramentas que possibilitem a padronização semântica entre os sistemas envolvidos nesse processo, como a inclusão de tesouros e ontologias.

O princípio FAIR Reutilizável (*Reusable*) denota a capacidade de adaptar os dados para que os pesquisadores possam reutilizá-los. Esse processo viabiliza que os dados sejam aproveitados pela comunidade científica e empregados em novos propósitos de pesquisa. No entanto, é importante estabelecer políticas que assegurem a acessibilidade tanto dos dados quanto dos metadados, visando a possíveis novos usos.

A aplicação eficaz dos Princípios FAIR pode conduzir a uma gestão mais organizada, facilitando a administração de conjuntos de dados pelos pesquisadores. Conforme enfatizado por Wilkinson *et al.* (2016), uma gestão adequada dos dados de pesquisa deixou de ser apenas um objetivo para se tornar uma pré-condição que sustenta a proliferação de novas descobertas científicas.

3.1.1 **FairDataBR: Avaliando a Aderência de Conjuntos de Dados aos Princípios FAIR**

Um desafio ao trabalhar com os princípios FAIR está relacionado com a necessidade de avaliar a aderência dos conjuntos de dados disponíveis a esses princípios. Nesse sentido, o grupo de pesquisa WRCO (*Web, Representação do Conhecimento e Ontologias*), cadastrado no Diretório dos Grupos de Pesquisa do Brasil do CNPq, desenvolveu uma ferramenta com o objetivo de avaliar em que medida um conjunto de dados adere aos princípios FAIR. A desenvolvida ferramenta foi intitulada de *FairDataBR* e está disponível na seguinte URL: <https://wrco.ufpb.br/fair/> (Figura 8). Para o desenvolvimento da ferramenta, foi empregada a linguagem de programação *JavaScript*. O código-fonte está disponível gratuitamente para qualquer pessoa interessada.

Figura 8 - Os Princípios orientadores FAIR



Fonte: FairDataBR(2024)

A ferramenta mencionada apresenta uma série de questões de forma objetiva, organizadas em blocos correspondentes aos respectivos princípios FAIR (Figura 9). Após o preenchimento dessas questões, a ferramenta calcula uma nota de zero (0) a dez (10) para cada princípio FAIR, além de calcular a média aritmética dos valores atribuídos a cada um deles. Com base nos resultados atribuídos a cada princípio, o usuário da ferramenta pode intervir no ambiente relacionado ao conjunto de dados avaliado para aprimorar qualquer aspecto que não esteja considerado adequado.

Figura 9 - Os princípios orientadores FAIR

R4. Os conjuntos de (meta)dados possuem proveniência detalhada? 1 6 *

Sim
 Não

R5. Os conjuntos de (meta)dados são acompanhados ou vinculados a uma descrição da origem do fluxo de trabalho que produziu os (meta)dados? 1 5 6 *

Rastreamento de Proveniência
 Declaração de Proveniência
 Sem Informações de Proveniência

R6. Os (meta)dados estão de acordo com padrões relevantes para o Domínio? 1 4 6 *

Padrões de Domínio
 Padrões Mínimos do Domínio
 Padrões Genéricos
 Nenhum Padrão Utilizado

Calcular Pontuação Salvar/Imprimir Avaliação

Pontuação Calculada

Aproveite e verifique como ficará sua pontuação em FairDataBR*

Copiar respostas para FairDataBR*

7.60	4.67	8.00	10.00	7.57
FINDABLE	ACCESSIBLE	INTEROPERABLE	REUSABLE	MÉDIA

Seção 1 FINDABLE:	Seção 2 ACCESSIBLE:	Seção 3 INTEROPERABLE:	Seção 4 REUSABLE:	MÉDIA FAIR:
7.60	4.67	8.00	10.00	7.57

Fonte: FairDataBR(2024)

Além das funcionalidades mencionadas, a ferramenta *FairDataBR* demonstrou ser bastante adequada no ensino dos princípios FAIR em turmas de cursos de graduação e pós-graduação.

3.2 OS PRINCÍPIOS CARE APLICADOS À GOVERNANÇA DE DADOS ÍNDIGENAS

Considerando a necessidade de uma Gestão de Dados de Pesquisa como forma de gerenciar os dados durante a investigação,

o tratamento dos dados pode levar em conta os Princípios CARE para Governança de Povos Indígenas, além de sua capacidade de preservação digital e compartilhamento de dados.

Os Princípios CARE¹¹ foram definidos durante a Semana Internacional de Dados e no evento co-organizado pela Plenária da *Research Data Alliance* “Princípios de Soberania de Dados Indígenas para a Governança do Workshop de Dados Indígenas”, 8 de novembro de 2018, Gaborone, Botsuana (Carrol *et al*, 2020).

No contexto da ciência aberta e dos dados abertos, tais princípios estão diretamente ligados aos povos indígenas e seus direitos e interesses, e se constituem como um conjunto de procedimentos que visam auxiliar os povos indígenas no tratamento, metadados, compartilhamento, uso e reusos dos dados abertos. Os Princípios CARE para Governança de Povos Indígenas, refletem sua atuação no contexto da inovação e governança dos dados.

Para Carrol *et al* (2020, p.3, tradução nossa)¹², os Princípios CARE para Governança de Povos Indígenas

capacitam os Povos Indígenas, mudando o foco da consulta regulamentada para relacionamentos baseados em valores que posicionam as abordagens de dados dentro das culturas indígenas e sistemas de conhecimento em benefício dos Povos Indígenas [...] Em última análise, esta mudança promove a participação equitativa nos processos de reutilização de dados, o que resultará em práticas mais equitativas.

¹¹ Disponível em: <https://www.gida-global.org/care>. Acesso em: 25 mar 2024.

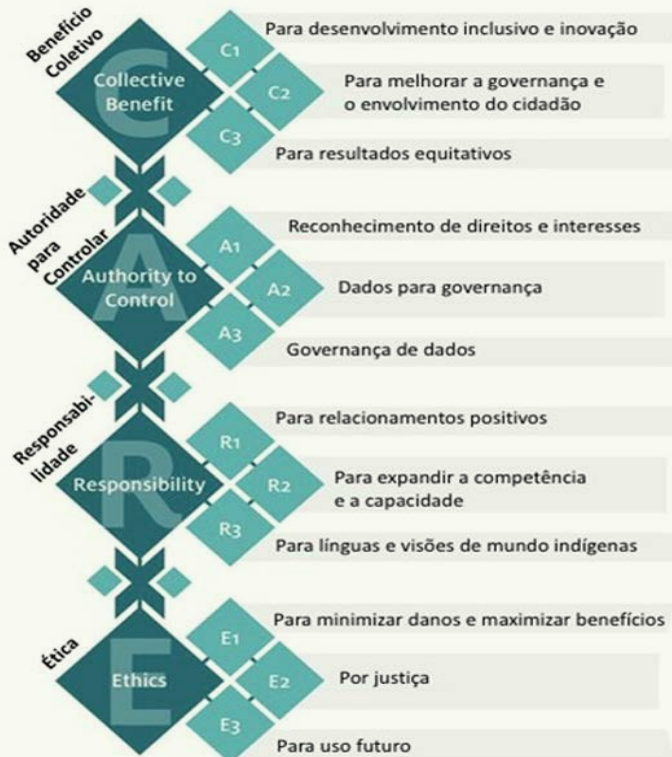
¹² No original: “empower Indigenous Peoples by shifting the focus from regulated consultation to value-based relationships that position data approaches within Indigenous cultures and knowledge systems to the benefit of Indigenous Peoples (Castellano 2004; Anderson *et al*. 2003). This shift ultimately promotes equitable participation in processes of data reuse, which will result in more equitable outcomes”.

4



Os Princípios CARE para Governança de Povos Indígenas são compostos por *Collective Benefit* /Benefício Coletivo, *Responsibility*/ Responsabilidade e *Ethics*/Ética, conforme expresso na Figura 10, com base na contribuição de Vidotti, Torino e Coneglian (2021, p. 214).

Figura 10 - Os Princípios CARE para Governança de Povos Indígenas



Fonte: Vidotti, Torino e Coneglian (2021, p. 214)

Os Princípios CARE para Governança de Povos Indígenas, complementam a abordagem centrada em dados existente representada nos “Princípios Orientadores FAIR para gestão e administração de dados

científicos” (Localizável, Acessível, Interoperável, Reutilizável). A Figura 11 mostra essa relação de complementaridade sobre os Princípios CARE para Governança de Povos Indígenas e os Princípios FAIR.

Figura 11 - Mescla dos Princípios FAIR e CARE

#BeFAIRandCARE



FAIR Guiding Principles for scientific data management and stewardship can be accessed [here](#)

Fonte: CARE Principles for Indigenous Data Governance (2024)

A comunidade indígena engloba mais de 90 países com mais de 370 milhões de indígenas, o que vale incidir que vivenciam mais de 5.000 de culturas diversas (Carrol *et al*, 2020). Para Carrol *et al* (2020), se faz necessário o entendimento dessa ecologia dos dados como modo de enfrentamento aos desafios na produção, uso, reuso e compartilhamento de dados que possam dimensionar as comunidades indígenas, bem como seus interesses e necessidades de governança.

CAPÍTULO 4

4 PLANO DE GESTÃO DE DADOS

De modo geral, podemos compreender que um **Plano de Gestão de Dados (PGD)**, conhecido como **Data Management Plan (DMP)** em Língua inglesa, é um documento que está associado a um conjunto de informações que delineiam um roteiro que contribui para o processo de gestão dos dados de pesquisa. A função primordial do PGD é auxiliar os pesquisadores no registro dos processos de coleta, tratamento e descrição dos dados de pesquisa, além de fornecer diretrizes essenciais para o compartilhamento, uso e reuso desses dados. O PGD contribui para a realização de uma curadoria efetiva dos dados de pesquisa.

O PGD é um instrumento estabelecido, frequentemente solicitado por várias universidades, agências de fomento e centros de pesquisa. Essas instituições estão implementando medidas para assegurar que os dados provenientes das pesquisas sejam gerenciados e preservados de forma adequada ao longo de todo o processo da investigação científica. De acordo com Veiga *et al* (2022), no seu Plano de Gestão de Dados De Pesquisa - PGD: Guia De Elaboração, são apontados alguns preceitos que devem ser levados em conta para a construção desse instrumento, quais sejam:

- Quais dados serão coletados ou criados? (Veiga *et al*, 2022).
- De que forma os dados serão coletados ou produzidos e como poderão ser reusados? (Veiga *et al*, 2022).
- Qual padrão de metadados acompanhará os conjuntos de dados? (Veiga *et al*, 2022).

- Que tipos de documentação acompanharão os conjuntos de dados? (Veiga *et al*, 2022).
- Como será realizada a gestão de questões éticas e os códigos de conduta? (Veiga *et al*, 2022).
- Como as questões legais de Direitos de Propriedade Intelectual (DPIs) e autoria serão gerenciadas? Que legislação é aplicável? (Veiga *et al*, 2022).
- Como serão feitos os *backups* e o armazenamento dos conjuntos de dados e dos metadados durante e após o processo de pesquisa? (Veiga *et al*, 2022).
- Como serão gerenciados o acesso e a segurança dos dados? (Veiga *et al*, 2022).
- Quais dados serão selecionados para retenção, compartilhamento e/ou preservação a longo prazo? (Veiga *et al*, 2022).
- Qual é o planejamento para a preservação a longo prazo dos conjuntos de dados? (Veiga *et al*, 2022).
- Como e quando os dados serão compartilhados?
- Existem possíveis restrições ao compartilhamento do conjunto de dados? (Veiga *et al*, 2022).
- Quem serão os responsáveis pela gestão dos dados? (Veiga *et al*, 2022).
- Que recursos serão requeridos para a entrega do plano? (Veiga *et al*, 2022).

Para a elaboração do PGD, os autores ressaltam a utilização dos Princípios FAIR em toda sua construção e servem como bússolas para os pesquisadores estruturarem seu material. Ao final, é recomendável a realização do plano em um repositório estruturado e confiável, de

preferência, da instituição ao qual o pesquisador esteja vinculado (Veiga et al, 2022). Embora seja uma prática recente no país, algumas agências de fomento e Instituições de Ensino Superior (IES) começaram a estabelecer políticas mandatórias para a gestão de dados de pesquisa, descrevendo todas as diretrizes necessárias para sua efetivação.

Uma das principais propostas do PGD é garantir que os dados sejam compreensíveis para outros pesquisadores interessados no mesmo conjunto de dados. Por isso, é importante que as diretrizes sejam detalhadas e que informações adicionais e contextualização sejam comunicadas de forma eficaz, junto com as licenças de uso.

As premissas estabelecidas para a inserção do PGD na gestão de dados descrevem as metodologias adequadas e outros atributos necessários para sua realização. A elaboração do PGD é um procedimento versátil, que pode exigir habilidades tecnológicas mais específicas para trabalhar e manipular os dados de pesquisa.

É importante pontuar que os dados de pesquisa tendem a ser diversos e não uniformes, com formatos variados que os tornam heterogêneos. Isso pode tornar a gestão de dados mais trabalhosa, mas quando bem gerenciados, os dados podem conduzir a pesquisas de maior qualidade e visibilidade (Choudhury et al, 2018).

A seguir, listamos alguns atributos que o PGD deve conter (Figura 12).

Figura 12 - Premissas para o Plano de Gestão de Dados

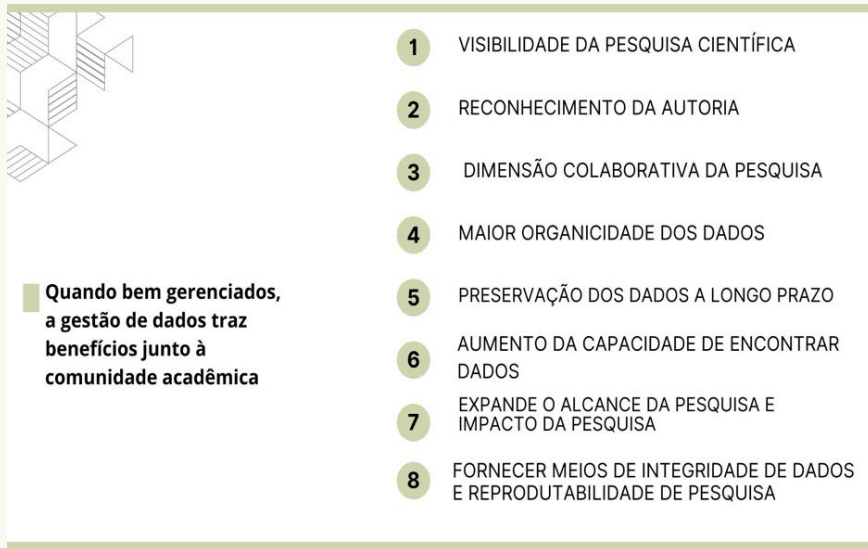


Fonte: Elaboração dos autores (2023)

Deste modo, é importante considerar que a gestão de dados não abrange todo o processo de gerenciamento dos dados científicos. Em vez disso, ela serve como uma bússola, fornecendo direção na gestão de grandes volumes de dados de pesquisa. Fatores como a adoção de um esquema de metadados e o armazenamento adequado são essenciais, especialmente considerando a importância dos dados no contexto da *e-Science*.

Sem uma descrição minuciosa do contexto tecnológico dos arquivos de dados, do contexto no qual os dados foram criados ou coletados, das medidas que foram feitas, dos detalhes espaciais e temporais, dos instrumentos usados, dos parâmetros e unidades e da qualidade dos dados e da sua proveniência, é improvável que os dados possam ser descobertos, interpretados, gerenciados e efetivamente usados e reusados. Os metadados cumprem essa tarefa, porque eles são a documentação dos dados (Sayão; Sales, 2015, p. 20).

Além disso, listamos alguns benefícios (Figura 13) que se tem a partir desta adoção, quais sejam:

Figura 13 - Principais benefícios que integram o PGD

Fonte: Elaboração dos autores (2023)

Uma gestão de dados eficaz pode direcionar o desenvolvimento e o avanço das pesquisas científicas, o que, quando bem gerenciado, pode levar a uma maior qualidade e visibilidade, trazendo benefícios consequentes para a comunidade científica. Embora a adoção dos Princípios FAIR possa garantir que esses dados sejam encontráveis, acessíveis, interoperáveis e reutilizáveis, essas premissas tornam-se fundamentais durante a fase de pesquisa ativa. Como Jones, Prior & White (2013, p. 5) destacam: “para fornecer suporte efetivo à gestão e ao compartilhamento de dados, uma instituição precisa de uma estratégia coerente e de um conjunto de serviços”.

As iniciativas de criação do PGD) ainda estão em estágio inicial em algumas instituições, agências de fomento e outras organizações. No entanto, parte dessas instituições em países como Reino Unido, Austrália, Holanda e Grã-Bretanha já exigem que os pesquisadores disponibilizem os dados publicamente ou forneçam seus próprios

Planos de Gestão de Dados. No Brasil, uma iniciativa notável tem sido da Fundação de Amparo à Pesquisa do Estado de São Paulo (Fapesp), que reconheceu a relevância dos dados de pesquisa para a comunidade científica e estabeleceu algumas regras para essa prática. A FAPESP entende que: “a gestão apropriada desses dados facilita a reprodutibilidade da pesquisa e permite promover novas pesquisas, graças à possibilidade de reuso e compartilhamento” (Fapesp, 2023). Nesse sentido, são estabelecidos alguns atributos essenciais que devem compor o PGD, a saber:

- a. Descrição dos dados e metadados produzidos pelo projeto - por exemplo, amostras, registros de coleta, formulários, modelos, resultados experimentais, *software*, gráficos, mapas, vídeos, planilhas, gravações de áudio, bancos de dados, material didático, entre outros (Fapesp, 2023, *online*);
- b. Quando aplicável, restrições legais ou éticas para compartilhamento desses dados, bem como políticas para garantir privacidade, confidencialidade, segurança, propriedade intelectual, entre outras (Fapesp, 2023, *online*);
- c. Política de preservação e compartilhamento - por exemplo, compartilhamento imediato ou somente após a aceitação da publicação associada. Deve ser especificado o período de embargo (antes do compartilhamento) e o período durante o qual os dados serão preservados e disponibilizados (Fapesp, 2023, *online*);
- d. Descrição de mecanismos, formatos e padrões para armazenar tais itens de maneira a torná-los acessíveis por terceiros. Essa descrição pode incluir o uso de repositórios e serviços de outras instituições (Fapesp, 2023, *online*).



De maneira similar, Sales e Sayão (2022, p.3) sugerem que a gestão do PGD envolve uma combinação de funções, todas voltadas para a adequação dos dados dentro de uma infraestrutura que forneça suporte aos pesquisadores. Guiados pelos Princípios FAIR, os autores propõem o que chamam de “serviços de gestão de dados de pesquisa”, que definem como “o conjunto de serviços informacionais, computacionais, científicos e administrativos oferecidos no âmbito da gestão de dados de pesquisa. Esses serviços são ancorados nas necessidades específicas das comunidades acadêmicas e científicas e têm como objetivo tornar os dados encontráveis, acessíveis, interoperáveis e reutilizáveis”. Portanto, esses serviços são divididos em quatro dimensões:

1. a governança, onde são discutidos os princípios norteadores do projeto de serviços de gestão de dados;
2. as infraestruturas técnicas, onde se incluem também as categorias de expertises necessárias;
3. os serviços informacionais, computacionais, científicos e administrativos;
4. os resultados da efetivação desses serviços manifestados pela FAIRificação dos dados (Sales; Sayão, 2022, p.4)

A implementação efetiva do PGD requer uma abordagem colaborativa que envolve tanto as instituições quanto a comunidade acadêmica. É importante levar em consideração que novas pesquisas exigem a adoção de metodologias inovadoras, estratégias de gestão de dados e práticas atualizadas.

Nesse contexto, as instituições têm o dever de prover treinamento eficaz aos pesquisadores, de modo a capacitá-los para a adequada gestão de dados. Além disso, elas devem oferecer consultoria para

a resolução de potenciais problemas e realizar *workshops* contínuos, que possam acompanhar a evolução e atualização das ferramentas.

Outro desafio estratégico se refere à manutenção da curadoria de dados em toda a instituição, o que deve ser feito em conformidade com as suas especificidades. Portanto, não existe um modelo padrão de PGD. Cada instituição precisa elaborar o seu próprio modelo, levando em consideração os recursos disponíveis, seus projetos de pesquisa e a sua visão estratégica.

CAPÍTULO 5

5 POLÍTICA DE GESTÃO DE DADOS DE PESQUISA

O estabelecimento de uma política de gestão de dados de pesquisa é uma etapa fundamental para qualquer instituição que tenha o objetivo de implantar um repositório de dados. Nesse sentido, a Fundação Oswaldo Cruz (2020) traz que essa etapa é necessária como parte das condições essenciais para o desenvolvimento de repositórios em longo prazo. A implementação de políticas de gestão de dados, bem como o monitoramento e a avaliação do uso dos dados depositados pelos pesquisadores, será conduzida por um Comitê Gestor.

O Comitê Gestor de um repositório deve ser composto por um(a) Coordenador(a) Geral, um(a) Coordenador(a) Técnico(a), o(a) Encarregado(a) de Dados (*Data Protection Officer - DPO*), um(a) Bibliotecário(a), e um(a) Representante da comunidade de usuário(as), da seguinte forma:

1. **Coordenador(a) geral:** Este(a) profissional lidera o Comitê Gestor e representa o repositório no contexto da instituição, bem como em outras instituições e entidades relacionadas à gestão de dados. O(a) Coordenador(a) geral deve ser, preferencialmente membro da instituição a qual o repositório de dados está vinculado;
2. **Coordenador(a) técnico(a):** Este(a) profissional é responsável por supervisionar as atividades técnicas do repositório, incluindo a manutenção da infraestrutura, o desenvolvimento de produtos de *software*, e a elaboração de políticas e orientações sobre o depósito e compartilhamento

de dados. A função de Coordenador(a) técnico(a), poderá ser exercida por qualquer pessoa (servidor ou não da instituição a qual o repositório de dados está vinculado) com as devidas habilidades;

3. **Encarregado(a) de Dados (*Data Protection Officer - DPO*):** Este(a) profissional é responsável por garantir a proteção dos dados armazenados no repositório. O(a) DPO assegura que o repositório esteja em conformidade com as leis e regulamentações de proteção de dados, incluindo a Lei Geral de Proteção de Dados (LGPD). Também contribui para o desenvolvimento de políticas de privacidade e segurança da informação para minimizar os riscos de vazamento ou uso indevido dos dados armazenados;
4. **Bibliotecário(a):** Este(a) profissional é responsável por organizar, armazenar e disponibilizar a produção científica da instituição, que inclui os dados gerados por pesquisas. Contribui para a adoção de boas práticas de gestão de dados e assegura que o repositório esteja em conformidade com as políticas e normas nacionais e internacionais de acesso aberto e preservação de dados de pesquisa. Auxilia os pesquisadores na elaboração do Plano de Gestão de Dados, no depósito de dados no repositório e na atribuição de identificadores persistentes para conjuntos de dados;
5. **Representante da comunidade de usuário(as):** Este(a) profissional representa a comunidade científica e acadêmica que faz uso do repositório de dados. É responsável por apresentar sugestões e críticas sobre o repositório ao Comitê Gestor.

Em síntese, a constituição de um Comitê Gestor bem estruturado e a implementação de políticas de gestão de dados são pilares

fundamentais para o estabelecimento e a sustentabilidade de um repositório de dados de pesquisa. É através da atuação coordenada dos diversos profissionais envolvidos, desde o Coordenador Geral até o Representante da comunidade de usuários, que se promove a segurança, acessibilidade e integridade dos dados depositados.

5.1 FERRAMENTAS ESPECIALIZADAS SUGERIDAS NO CONTEXTO DO REPOSITÓRIO DE DADOS

Recomendamos o uso de ferramentas de *software* especializadas, como *DMPtool* ou o *DMPonline*¹³, para a elaboração do PGD correspondente ao conjunto de dados que será depositado no repositório pelo seu respectivo detentor. Essas ferramentas foram especificamente projetadas para auxiliar os pesquisadores na elaboração de PGDs e garantir a conformidade com as políticas de gestão de dados associadas ao repositório. Sugerimos que todos os conjuntos de dados depositados no repositório de dados devem estar associados a um PGD.

O *DMPtool* é uma ferramenta *online* que disponibiliza modelos de PGD personalizáveis de acordo com a área de pesquisa, financiador e instituição. O *DMPtool* também fornece orientações e dicas para ajudar os pesquisadores a elaborar um PGD eficiente e completo. Já o *DMPonline* é outra ferramenta *online* que, além de oferecer modelos de PGD personalizáveis, permite que os usuários criem um PGD desde seu início. Adicionalmente, o *DMPonline* disponibiliza recursos para revisão e comentários do PGD por outros membros da equipe de pesquisa ou por um avaliador externo.

13 O *DMPtool* e o *DMPonline* podem ser acessados respectivamente através das seguintes URLs: <https://dmptool.org/> e <https://dmponline.dcc.ac.uk/>

Ao utilizar a *DMPtool* ou o *DMPonline* na elaboração de um PGD, os pesquisadores podem assegurar que o plano seja abrangente e esteja em conformidade com as políticas de gestão de dados associadas com o repositório. Isso eleva a qualidade e a confiabilidade dos dados depositados no repositório de dados, bem como a possibilidade de reutilização por outros pesquisadores.

5.2 ESPECIFICAÇÕES PARA O REPOSITÓRIO DE DADOS

Portanto, levando em consideração as particularidades de cada repositório de dados, as políticas para a gestão de dados precisam ser dinâmicas e voltadas para a ciência aberta. Dessa forma, o conjunto de diretrizes exige a criação de espaços de diálogo dentro da comunidade científica capazes de transmitir os aspectos obrigatórios que regerão os dados de pesquisas.

Nesse sentido, sugerimos um modelo de PGD que engloba os elementos como: Responsabilidades e autoridade, Coleta, Armazenamento de dados, Qualidade dos dados, Acessibilidade, Licenciamento, Integridade, Privacidade, Responsabilidade, Transparência, Aspectos Éticos e Período de embargo.

- **Responsabilidades e autoridade:** foca na responsabilidade pela gestão do repositório de dados, enquanto os usuários que depositam dados e materiais são responsáveis por garantir que os dados e materiais depositados estejam em conformidade com as políticas de gestão de dados específicas do repositório;
- **Coleta e armazenamento de dados:** o repositório de dados deve possibilitar que os usuários depositem dados e materiais de pesquisa em vários formatos, incluindo conjuntos

de dados, *software*, imagens, vídeos e apresentações. O repositório de dados armazena esses dados e os disponibiliza para *download*;

- **Qualidade dos dados:** a priori o repositório de dados não faz validações dos dados e materiais depositados. É responsabilidade dos usuários garantir que os dados e materiais estejam completos, precisos e consistentes. Eventualmente, sugestões para garantir a aderência dos conjuntos de depositados aos padrões internacionais que garantam o uso e reúso desses poderão ser realizados pelos gestores do repositório;
- **Acessibilidade:** o repositório de dados deve garantir que os dados depositados sejam acessíveis ao público de forma aberta e gratuita, sem restrições ou barreiras. Além disso, deverá ser especificado neste tópico, o período para que os dados estarão disponíveis para a comunidade acadêmica e demais interessados, como por exemplo: 1 ano, 5 anos ou liberação sem datas;
- **Licenciamento:** é necessário definir as condições de uso dos dados de maneira clara e explícita, para que os usuários possam compreender e seguir as regras de uso. Uma possibilidade seria a adoção da Licença *Creative Commons* (CC) para licenciar os dados depositados por seus usuários. Isso significa que os usuários que depositam seus dados no repositório de dados podem escolher a licença CC que melhor se adequa às suas necessidades, permitindo que outros usem e compartilhem seus dados sob as condições estabelecidas na licença. Em repositórios de dados abertos é comum utilização da licença CC-BY (Atribuição) para o depósito de conjuntos de dados. A licença CC-BY é uma das mais frequentes em repositórios de dados de pesquisa,

pois autoriza os usuários a reutilizar, modificar e distribuir os dados, inclusive para fins comerciais, desde que atribuam crédito ao autor original. Contudo, é fundamental salientar que a escolha da licença CC-BY deve ser ponderada de acordo com cada situação, levando em conta a natureza dos dados e o propósito de sua coleta;

- **Integridade:** é importante assegurar que os dados sejam precisos, confiáveis e autênticos, e que as alterações ou correções sejam registradas e documentadas;
- **Armazenamento:** é preciso definir as políticas de armazenamento e preservação dos dados, incluindo medidas de segurança e *backup*, para garantir a disponibilidade dos dados a longo prazo;
- **Privacidade:** deve-se garantir que os dados pessoais sejam tratados de forma adequada e respeitando a privacidade dos indivíduos;
- **Transparência:** é necessário fornecer informações claras e transparentes sobre as políticas de gestão de dados do repositório de dados, incluindo a política de privacidade e as condições de uso dos dados;
- **Aspectos Éticos e Legais:** os pesquisadores detentores de conjuntos de dados devem informar aos usuários sobre como fazer uso de dados e como obter consentimento para o uso futuro desses materiais, como por exemplo, ao utilizá-los, citar a origem dos dados;
- **Período de embargo:** o período de embargo é um intervalo preestabelecido durante o qual os dados permanecem inacessíveis para o público. Geralmente, essa restrição é definida pelo detentor dos dados (Sant’ana, 2019). Essa especificidade deve estar explícita na política de gestão

dos dados. O período de embargo pode ser necessário para permitir que os pesquisadores publiquem seus resultados antes da disponibilização dos dados subjacentes a terceiros, colaborando assim para a proteção da propriedade intelectual, para assegurar a privacidade dos participantes da pesquisa ou para atender a outras obrigações legais ou éticas. Após o término do período de embargo, os dados podem ser tornados acessíveis ao público em geral, sujeitos a quaisquer restrições adicionais que possam ser estabelecidas pelo detentor dos dados ou pelo próprio repositório de dados.

5.3 PRIVACIDADE E SEGURANÇA EM REPOSITÓRIOS DE DADOS

A privacidade e a segurança dos indivíduos são valores fundamentais e, portanto, os repositórios de dados não deveriam possibilitar a publicação de dados que possam violá-las. A Lei Geral de Proteção de Dados Pessoais (LGPD), Lei nº 13.709/2018, estabelece os requisitos necessários para a proteção dos dados pessoais. Isso implica que todos os dados que permitem a identificação de um indivíduo devem ser manuseados com cuidado e devidamente protegidos. Os repositórios de dados devem aderir às diretrizes estabelecidas pela LGPD para garantir a proteção da privacidade dos indivíduos.

Alguns conjuntos de dados podem conter informações sensíveis que, se divulgadas, podem comprometer a privacidade e segurança dos indivíduos envolvidos. Nessas situações, os detentores dos dados são responsáveis por anonimizá-los antes de disponibilizá-los no repositório de dados. A anonimização é um processo que transforma

os dados originais em dados que não permitem a identificação pessoal, garantindo que eles possam ser utilizados para fins de pesquisa, sem comprometer a privacidade dos indivíduos envolvidos.

5.4 ATRIBUIÇÃO DE IDENTIFICADORES PERSISTENTES

O crescente volume de dados gerados em diversos campos do conhecimento demanda uma infraestrutura robusta e eficiente para seu respectivo armazenamento, compartilhamento e recuperação. Nesse diapasão, os repositórios de dados desempenham um papel fundamental como intermediários entre produtores e suas comunidades de usuários. A atribuição de identificadores persistentes para conjuntos de dados depositados em repositórios é um aspecto chave nesse processo, pois facilita o acesso e a identificação das associações relacionadas a esses dados.

Um identificador persistente é uma designação única e constante associada a um objeto digital ou conjunto de dados, que possibilita encontrá-lo e recuperá-lo consistentemente ao longo do tempo. O *Digital Object Identifier* (DOI)¹⁴ é um exemplo de identificador persistente, amplamente utilizado em publicações científicas, conjuntos de dados e outros objetos digitais. Os DOIs são gerenciados pela *International DOI Foundation* (IDF) (Figura 14) e garantem interoperabilidade, persistência e resolução de identificadores.

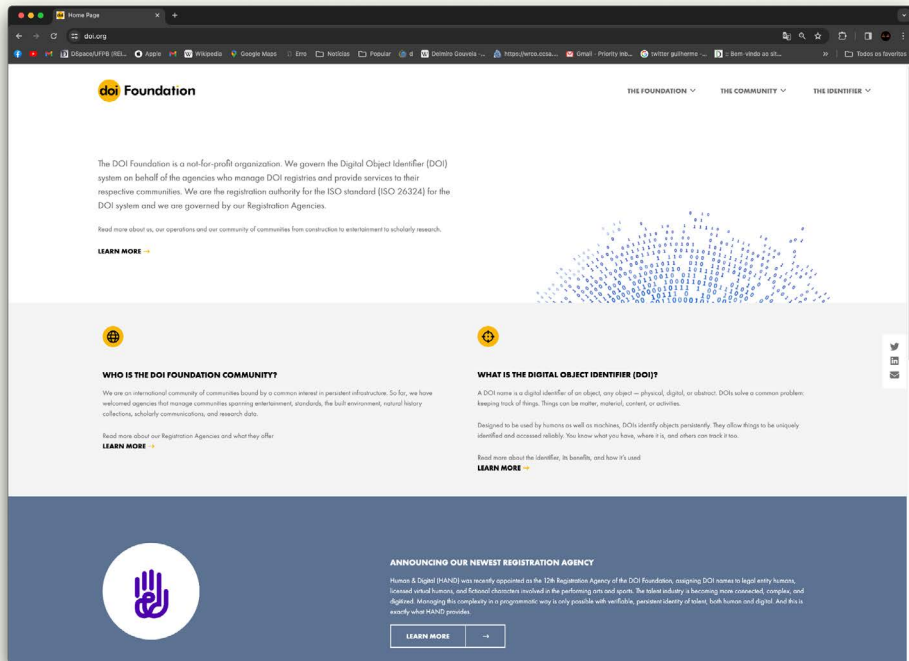
No caso específico dos DOI atribuídos a conjuntos de dados de dados, esses podem ser fornecidos pela instituição *DataCite*. Através dos serviços do *DataCite*, é possível fornecer suporte à criação e gestão de DOIs, bem como ao registro de metadados. As ferramentas disponibilizadas possibilitam o aprimoramento dos fluxos de trabalho

¹⁴ Disponível em: <https://www.doi.org/> Acesso em: 02 mai 2023.



de pesquisa por meio da integração de serviços que contribuem para a descoberta e reutilização de recursos (Datacite, 2024).

Figura 14 - Home do *Digital Object Identifier (DOI)*



Fonte: DOI (2024)

Os repositórios de dados devem atribuir um DOI específico para cada conjunto depositado. Esse DOI identifica inequivocamente o conjunto de dados em relação ao repositório, facilitando seu acesso e a identificação das associações relacionadas. A concessão de DOIs no repositório devem seguir as diretrizes e padrões estabelecidos pela IDF, garantindo a conformidade e a interoperabilidade com outros sistemas de identificação, apresentando benefícios tanto para os produtores quanto para os usuários de dados. Para os produtores, os DOIs oferecem

uma maneira eficaz de obter crédito e reconhecimento pelo trabalho realizado, já que podem ser facilmente citados e rastreados. Para os usuários, os DOIs facilitam o acesso aos conjuntos de dados, permitindo a localização e a recuperação das informações de maneira rápida e eficiente.

CAPÍTULO 6

6 CONSIDERAÇÕES FINAIS

A Gestão de Dados de Pesquisa é uma atividade multifacetada que engloba pesquisadores de diversas áreas do conhecimento e utiliza várias TDICs. Considerando essa realidade e nossa experiência prévia com pesquisadores diretamente envolvidos nos processos de coleta e descrição de dados, percebemos que os principais desafios para a efetiva consolidação da Gestão de Dados de Pesquisa na esfera científica não estão, em sua maioria, relacionados a aspectos tecnológicos. Pelo contrário, os maiores obstáculos residem no treinamento dos pesquisadores e na conscientização dos gestores das instituições de ensino e pesquisa sobre a importância dessa atividade.

Durante a realização de *workshops* com pesquisadores de diversas instituições de ensino e pesquisa na área da biodiversidade, identificamos uma lacuna significativa em sua compreensão de uma série de tópicos, tais como: Ciclos de Vida de Dados (CVDs), Planos de Gestão de Dados (PGDs), padrões de metadados, citação de dados, licenciamento de dados, repositórios de dados, períodos de embargo, dados abertos, Princípios FAIR, Princípios CARE, entre outros. Desta forma, ressaltamos que a educação abrangente é importante em todas as dimensões relacionados à Gestão de Dados de Pesquisa e fundamental para o êxito dessa atividade. Além disso, é importante destacar que todos os pesquisadores com quem interagimos reconhecem a importância e o valor dos dados que coletam ou produzem.

No contexto das instituições de ensino e pesquisa às quais os pesquisadores estão afiliados, observamos que, frequentemente, os gestores dessas instituições não compreendem plenamente o

significado da Gestão de Dados de Pesquisa, nem sua importância. Nesse sentido, deparamo-nos com diversos desafios para implementar essa gestão de forma eficaz nas instituições, que incluem, entre outros aspectos: a escassez de pessoal especializado em tecnologia da informação para configurar e dar suporte aos repositórios de dados; a falta de familiaridade dos bibliotecários com a Gestão de Dados de Pesquisa, tornando necessária a presença de bibliotecários especializados em dados nas instituições; e as dificuldades diversas para obter identificadores persistentes dentro do âmbito institucional.

REFERÊNCIAS

BORGMAN, C. L. **The conundrum of sharing research data**. Journal of the American Society for Information Science and Technology, v. 63, n. 6, p. 1059-1078, 2012. Disponível em: <https://bit.ly/3L0c00V> 10.1002/asi.22634. Acesso em: 30 mar. 2023.

BORGMAN, Christine L. **Big data, little data, no data**. Scholarship in the networked world. Cambridge-London: The MIT Press, 2015. (Kindle version).

CARROLL, S. R. et al. The CARE Principles for Indigenous Data Governance. **Data Science Journal, France**, v. 19, n. 1, p. 43, nov. 2020. DOI: <http://doi.org/10.5334/dsj-2020-043>. Disponível em: <https://bit.ly/3MM1RGc>. Acesso em: 25 mar 2023

CARE PRINCIPLES FOR INDIGENOUS DATA GOVERNANCE . Disponível em: <https://www.gida-global.org/care>. Acesso em: 25 mar 2023

CAVALCANTI, M. T; SALES, L. F.; PIMENTA, R. M . Compartilhamento e Acesso a Dados de Pesquisa em Humanidades Digitais. **Cadernos De Biblioteconomia, Arquivística E Documentação**, v. 2018, p. 30-45, 2018. Disponível em: <https://cip.brapci.inf.br/autoridade/v/110022>Acesso em: 8 jan. 2024.

CHIWARE, E.; MATHE, Z. **Academic libraries' role in research data management services**: a South African perspective. South African Journal of Libraries and Information Science, v. 81, n. 2, p. 1-10, 2015.

CREATIVE COMMONS. **Atribuição 4.0 Internacional (CC BY 4.0)**. (2023). Disponível em: <https://bit.ly/3N3R7ol> Acesso em: 12 abr. 2023.

DATA CITE. **DataCite**. (2024). Disponível em: <https://datacite.org/>. Acesso em: 27 mar.. 2019.

DATAONE. Programa de Pesquisa em Biodiversidade. [S. l.]: **Dataone**, [2019]. Disponível em: <https://bit.ly/3UYDjxm> . Acesso em: 10 ago. 2019.

DIAS, G; ANJOS, R. L; A.; RODRIGUES, A. A. Dados científicos: as práticas de gestão dos pesquisadores brasileiros na Ciência da Informação. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 18., 2017, Marília (SP). **Anais** [...]. Marília: Unesp, 2017. Disponível em: <https://bit.ly/3XV3GEb>. Acesso em: 8 jan. 2024.

DIGITAL CURATION CENTRE. **Checklist for a Data Management Plan. v.4.0**. Edinburgh: Digital Curation Centre, 2013. Disponível em: http://www.dcc.ac.uk/sites/default/files/documents/resource/DMP/DMP_Checklist_2013.pdf Acesso em: 20 mar. 2024.

DUDZIAK, E. **Competências do bibliotecário na gestão de dados de pesquisa, comunicação científica e acesso aberto**. 2016. Disponível em: <https://bit.ly/3y6WYAJ>. Acesso em: 17 maio 2018.

ERWAY, R. **Starting the conversation: university-wide research data management policy**. Dublin, OH: OCLC Research, 2013. Disponível em: <https://www.oclc.org/content/dam/research/publications/library/2013/2013-08.pdf> Acesso em: 11 ago. 2019.

FAIRDATABR. **FairDataBR: uma ferramenta para a avaliação de conjuntos de dados**. Disponível em: <https://wrco.ufpb.br/fair/>. Acesso em 26 mar. 2024.

FOX, P.; HENDLER, J. E-science semântica: o significado codificado na próxima geração de ciência digitalmente aprimorada. In: HEY, T.; TANSLEY, S.; TOLLE, K. (Org.). **O quarto paradigma: descobertas científicas na era da e-science**. São Paulo: Oficina de Textos, 2011.

FUNDAÇÃO DE AMPARO À PESQUISA DO ESTADO DE SÃO PAULO. **PITE**. São Paulo: FAPESP, 21 mar. 2019. Disponível em: <http://www.fapesp.br/61>. Acesso em: 23 FEV. 2023.

FUNDAÇÃO OSWALDO CRUZ. Presidência. Vice-presidência de Ensino, Informação e Comunicação.: princípios e diretrizes. Rio de Janeiro, 2020. 19 p.

GRANT, R. Recordkeeping and research data management: a review of perspectives. **Records Management Journal**, v. 27, n. 2, p. 159-174, 2.

GRAY, J. Jim Gray on eScience: a transformed scientific method. In: HEY, T.; TANSLEY, S.; TOLLE, K. (Eds.). **The fourth paradigm: data intensive scientific discovery**.

HEY, T.; TANSLEY, S.; TOLLE, K. (ed.). **The fourth paradigm: data-intensive scientific discovery**. Washington: Microsoft Research, 2009.

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

JONES S., PRIOR G; WHITE, A. (2013). **How to develop research data management services – a guide for heis**. Edinburgh: Digital Curation Centre. Disponível em: <https://bit.ly/3xKhPJQ>. Acesso em: 23 fev. 2023.

JONES, S. Research data policy briefing. **United Kingdom: DCC**, nov. 2011. Disponível em: <https://bityli.com/6doBdg>. Acesso em: 10 mar. 2023.

LIPPINCOTT, J. K. Library and information technology support of e-science in the western context. In: MARCUM, D. B.; GEORGE, Gerald. **The data deluge: can libraries cope with e-science?** Santa Barbara, California: ABC CLIO, 2010.

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

OPEN KNOWLEDGE FOUNDATION. **Manual dos dados abertos: governo. Tradução e adaptação de Comunidade Transparência Hacker**. 2010. Disponível em: <https://bit.ly/3XXvvf1>. Acesso em: 25 maio 2019.

OECD Principles and Guidelines for Access to Research Data from Public Funding. Paris: **Organization for Economic Co-operation and Development**, 2007. Disponível em: <https://bit.ly/3UXZsfj>. Acesso em: 12 abr. 2023.

PRINCIPLES FAIR. **Go FAIR**. (2024) Disponível em: <https://bit.ly/2Qt21U0>. Acesso em: 20 fev. 2023.

RESEARCH DATA ALLIANCE. **Data Policy**. [S. l.]: Research Data Alliance, 2015. Disponível em: <https://bit.ly/3Lqwrpg>. Acesso em: 12 mar. 2023.

RODRIGUES, A. A. DIAS, G. A . Principais aspectos dos repositórios de dados de pesquisa no Brasil. **Bibliocanto**, v. 9, 2023. Disponível em: <https://periodicos.ufrn.br/bibliocanto/article/view/34166>. Acesso em: 07 nov. 2023.

RODRIGUEZ-IGLESIAS, A. et al. Publishing FAIR Data: an exemplar methodology utilizing PHI-Base. **Frontiers in plant science**. 7, 641. 2016. Disponível em <https://curtlink.com/nhA1toz>. Acesso em: 267 mar. 2024.

SALES, L. F. ; SAYÃO, L. F . Gestão de Dados de Pesquisa: apresentação. **LIINC EM REVISTA**, v. 15, p. 2-4, 2019.

SALES, L. F. ; SAYÃO, L. F. Proposta de modelo de serviço de gestão de dados de pesquisa. **ATOZ: NOVAS PRÁTICAS EM INFORMAÇÃO E CONHECIMENTO**, v. 11, p. 1-13, 2022.

SANT'ANA, R. C. G. Ciclo de vida dos dados: uma perspectiva a partir da ciência da informação. **Informação & Informação**, Londrina, v. 21, n. 2, p. 116-142, maio/ago. 2016. Disponível em: <https://ojs.uel.br/revistas/uel/index.php/informacao/article/view/27940/20124>. Acesso em: 26 mar. 2024.

SANT'ANA, R. C. G. Campo Informacional Resultante da Interação de Ciclos de Vida dos Dados. In: DIAS, G.; FREIRE, B. (orgs). **Dados Científicos: perspectivas e desafios**. João Pessoa: Editora UFPB. 2019.

SANTOS, P. L. V. A. C.; SANTANA, R. C. G. Transferência da informação: análise para valoração de unidades de conhecimento. **DataGramZero**, v. 3, n. 2, 2002. Disponível em: <https://brapci.inf.br/#/v/5349>. Acesso em: 26 mar. 2024.

SANTOS, P. L. V. A. C.; SANTANA, R. C. G. Dado e granularidade na perspectiva da informação e tecnologia: uma interpretação pela ciência da informação. **Ciência da Informação**, v. 42, n. 2, 2013. Disponível em: <https://brapci.inf.br/#/v/17173>. Acesso em: 26 mar. 2024.

SAYÃO, L. F.; SALES, L. F. Dados de pesquisa: contribuição para o estabelecimento de um modelo de curadoria digital para o país. **Pesquisa Brasileira em Ciência da Informação e Biblioteconomia**, v. 8, n. 2, 2013. Disponível em: <http://eprints.rclis.org/22562/>. Acesso em: 26 mar. 2024.

SAYÃO, L. F.; SALES, L. F. **Guia de gestão de dados de pesquisa para bibliotecários e pesquisadores**. Rio de Janeiro: Comissão Nacional de Energia Nuclear, 2013. Disponível em: <https://bit.ly/3XXukMN>. Acesso em: 14 nov. 2020.

SAYÃO, Luís Fernando ; SALES, L.F. Ciberinfraestrutura de Informação para a Pesquisa: proposta e Integração entre repositório institucional, repositório de dados e Cris. **Informação & Sociedade** (Ufpb. Online), V. 25, P. 163-184, 2015. Disponível em: <https://periodicos.ufpb.br/ojs2/index.php/ies/article/view/23998>. Acesso em: 24 jan. 2023.

SETZER, V. W. **Dado, Informação, Conhecimento e Competência**. (2015). Disponível em: <https://www.ime.usp.br/~vwsetzer/dado-info.html>. Acesso em: 26 mar. 2024.

SILVA, F. C. C. da. **Gestão de Dados Científicos**. Rio de Janeiro: Interciência, 2019. 146 p.

VEIGA, V. S. O.; DIB, S. ; HENNING, P.; NASCIMENTO, F. J. T. ; GUANAES, P. ; PENEDO, E. ; GUIMARÃES, C. . **Plano De Gestão de Dados de Pesquisa - PGD: Guia de Elaboração**. 1. ed. Rio de Janeiro: FIOCRUZ, ICICT, Rede de Bibliotecas, 2022.

VEIGA, V. S. de O.; SILVA, C. H.; BORGES, M. M. Modelo de fatores que influenciam no comportamento de compartilhamento de dados de pesquisa (MFDados). In: BORGES, M. M.; CASADO, E. S. (Org.). **Sob a lente da ciência aberta: olhares de Portugal, Espanha e Brasil**. Coimbra: Imprensa da Universidade de Coimbra, 2021. p. 153-187.

VEIGA, V. S. O.; DIB, S. ; HENNING, P.; NASCIMENTO, F. J. T. ; GUANAES, P. ; PENEDO, E. ; GUIMARÃES, C. . **Plano De Gestão de Dados de Pesquisa - PGD: Guia de Elaboração**. 1. ed. Rio de Janeiro: FIOCRUZ, ICICT, Rede de Bibliotecas, 2022.

VIDOTTI, S. A. B. G.; TORINO, E; CONEGLIAN, C. S.

#SejaJUSTOeCUIDADOSO: princípios FAIR e CARE na gestão de dados de pesquisa. In: SALES, L. F.; VEIGA, V. S.; HENNING, P.; SAYÃO, L. F. (org.). **Princípios FAIR aplicados à gestão de dados de pesquisa**. Rio de Janeiro: Ibict, 2021. p. 201-214. DOI: <https://doi.org/10.22477/9786589167242.cap15>. Disponível em: <https://ridi.ibict.br/handle/123456789/1182>. Acesso em: 25 mar 2024.

VOSS, A.; ASGARI-TARGHI, M.; PROCTER, R.; FERGUSSON, D.

Adoption of e-Infrastructure services: configurations of practice. **Philosophical Transactions of Royal Society**, v. 368, n. 1926, p. 4161-4176, 2 Aug. 2010. Disponível em: <https://bit.ly/3YVNFzd>. Acesso em: 5 mar. 2014.

WILKINSON, M. D. et al. The FAIR Guiding Principles for scientific data management and stewardship. **Scientific Data**. 2016. Disponível em: <https://go.nature.com/3loEPD3> . Acesso em: 07 nov. 2017.

WHYTE A, TEDDS J (2011) **Defending research data management**. Edinburgh: Digital Curation Center. Disponível em: <https://bit.ly/3IOM1d7>. Acesso em: 24 jan. 2023.

GLOSSÁRIO

Anonimização: é o processo que inviabiliza a identificação de uma pessoa a partir dos dados disponíveis.

Ciberinfraestrutura: termo que se refere a um conjunto de tecnologias da informação e comunicação, projetadas para propiciar novas abordagens metodológicas, formas de trabalho, pesquisas e geração de conhecimentos científicos.

Creative Commons: é uma organização sem fins lucrativos, apoiada por voluntários, que facilita o compartilhamento e uso de conteúdos criativos. Ela oferece uma variedade de licenças de direitos autorais que os criadores podem usar para especificar as condições sob as quais suas obras podem ser utilizadas por terceiros.

Ciclo de Vida dos Dados (CVD): Trata-se de um processo contínuo de gestão, planejamento, coleta, integração, descrição, controle de qualidade e descarte empreendida nos conjuntos de dados que pode ser aplicado em todas as etapas dos processos de investigação empreendidos.

Dado: “[...]elemento básico, formado por signo ou conjunto finito de signos que não contém, intrinsecamente, um componente semântico, mas somente elementos sintáticos.” (Santos e Sant’Ana, 2002, p.3),

Dados computacionais: São elementos gerados por modelos executados por computador, simulações ou fluxo de trabalho. São comuns tanto nas ciências sociais quanto nas humanidades. Esses podem ser reutilizados futuramente, desde que haja uma documentação extensiva do *hardware*, *software* e dos dados de entrada utilizados (Borgman, 2015).

Dados de pesquisa: São resultados oriundos de investigações científicas e podem assumir diversos formatos, tais como números, figuras, vídeos, *software*, áudios, imagens, imagens em movimento, gráficos e visualizações de dados, entre outros.

Dados do tipo registro: Qualquer registro de fenômeno ou atividade humana pode ser tratado como dado para pesquisa (Borgman, 2015).

Dados experimentais: São dados obtidos por meio de procedimentos controlados realizados em laboratórios para testar ou estabelecer hipóteses e descobrir ou confirmar novas leis. Esses dados podem vir de diversas áreas, como resultados de pesquisas em química realizadas em laboratórios de via úmida, experimentos de física em colisores lineares e estudos de psicologia conduzidos em laboratórios ou em campo (Borgman, 2015).

Dados observacionais: São caracterizados pela identificação, notificação e registro de fatos ou ocorrências de determinado fenômeno. Normalmente, eles são coletados com o auxílio de instrumentos, como *notebooks* ou satélites (Borgman, 2015).

Data Observation Network for Earth (DataONE): É um projeto comunitário que visa fornecer acesso a dados em repositórios e oferece auxílio para as pesquisas com dados ambientais e da Terra. Disponibiliza uma ciberinfraestrutura com vistas ao atendimento das pesquisas científicas, bem como oferece uma guia de gerenciamento e compartilhamento de dados, além de estabelecer o Ciclo de Vida dos Dados. Tem o suporte da *National Science Foundation (NSF)*.

Data steward: Profissional responsável pela gestão dos dados de pesquisa de maneira sustentável e de longo prazo, abrangendo todas as fases, desde o planejamento do estudo até a coleta, análise, armazenamento e compartilhamento dos dados. Inclui atividades essenciais para assegurar que os dados de pesquisa digitais sejam localizáveis, acessíveis, interoperáveis e reutilizáveis (FAIR), envolvendo o arquivamento e o reuso por terceiros.(Sales; Sayão, 2019).

Digital Curation Centre (DCC): É um centro que tem especialização em informação digital com o objetivo de promover capacitação, capacidade e habilidades em gerenciamento de dados de pesquisa. O centro oferece auxílio prático sobre questões que envolvam dados informacionais no tocante a armazenar, gerenciar, proteger e compartilhar dados de pesquisa, além de oferecer serviços de consultoria, políticas e planejamento de dados.

Dataverse Project: É um repositório de código aberto que tem o objetivo de compartilhar, citar, preservar, explorar e analisar dados de pesquisa, bem como facilitar a disponibilização de dados para as comunidades interessadas. Tem a capacidade de armazenar inúmeras coleções de dados, em que cada coleção disponível, contém metadados descritivos e arquivos de dados. Foi uma iniciativa da *Harvard University*.

DOI: *Digital Object Identifier* é um exemplo de identificador persistente, amplamente utilizado em publicações científicas, conjuntos de dados e outros objetos digitais. Esse sistema garante que esses recursos permaneçam acessíveis, mesmo que a URL mude ao longo do tempo.

e-Science: caracterizada pelo uso de instrumentos e aparatos tecnológicos que auxiliam na manipulação e análise dos grandes volumes de dados produzidos diariamente na ciência contemporânea. Essa abordagem, que surgiu no século XX, promove a colaboração e a multidisciplinaridade, sendo considerada o quarto paradigma da ciência, de acordo com Gray (2007).

Gestão de dados de pesquisa: se refere à organização de dados, desde sua coleta no início do ciclo de pesquisa até a disseminação e o arquivamento dos resultados significativos. Esta definição abrange um conjunto de práticas e processos distintos, envolvendo o design, criação, segurança, armazenamento, recuperação, preservação, reutilização e compartilhamento de dados. Tais processos consideram aspectos como capacidades tecnológicas, questões éticas, legislação e estrutura gerencial.

Lei Geral de Proteção de Dados Pessoais: define normas obrigatórias para a coleta, processamento e armazenamento de dados pessoais. Inspirada na Regulação Geral de Proteção de Dados (*General Data Protection Regulation - GDPR*), que foi implementada na União Europeia em 2018, esta legislação ocasionou significativos impactos tanto para empresas quanto para consumidores.

Período de embargo: período determinado durante o qual o acesso público aos dados é restrito, geralmente por decisão do detentor desses dados.

Plano de Gestão de Dados (PGD): conjunto de normas que estabelecem as diretrizes para a gestão dos dados de pesquisa. A principal função do PGD é orientar os pesquisadores na coleta e manipulação dos dados de pesquisa, fornecendo instruções essenciais para o compartilhamento, curadoria, uso, reuso e redistribuição desses dados.

Política de Gestão de Dados: Um conjunto de princípios que estabelecem um quadro orientador para a gestão de dados. Uma política de gestão de dados pode ser implementada para tratar de aspectos estratégicos, incluindo o acesso aos dados, questões legais pertinentes, administração e responsabilidades de custódia de dados, aquisição de dados e outras questões relevantes. (Research Data Alliance, 2015).

Princípios FAIR: são um acrônimo para *Findable* (Encontrável), *Accessible* (Acessível), *Interoperable* (Interoperável) e *Reusable* (Reutilizável). Eles recomendam que cada conjunto de dados deve ter um identificador global único, associado a metadados contextuais e pesquisáveis, tornando-os 'Encontráveis'. Os dados e metadados devem ser acessíveis

através de um protocolo aberto e padrão, tornando-os 'Acessíveis'. Eles devem empregar uma linguagem de representação formal, de ampla aplicabilidade, além de vocabulários e ontologias abertos e amplamente aceitos para o domínio relevante, tornando-os 'Interoperáveis'. Por fim, para serem 'Reutilizáveis', os dados devem ser ricamente descritos, com abundância de referências cruzadas e um mecanismo claramente definido para acessar informações de proveniência e licença (Rodriguez-Iglesias *et al*, 2016).

APÊNDICE A - LICENÇAS EM CAMADAS DO CREATIVE COMMONS



LICENÇAS CC EM TRÊS CAMADAS

As licenças do CC estão divididas em três camadas como forma de especificar de um modo adequado os usos e atribuições.

1. TEXTO LEGAL

As nossas licenças públicas de direito de autor e de direitos conexos incorporam um design único e inovador em “três camadas”. Cada licença começa por ser um instrumento legal tradicional, no gênero de linguagem e formato de texto que os advogados conhecem e adoram. Chamamos a esta camada de cada licença o Texto Legal.

2. LEGÍVEL POR HUMANOS

Mas como na realidade a maioria dos criadores, educadores e cientistas não são advogados, também disponibilizamos as licenças num formato que pode ser lido por todos – o Resumo Explicativo (também conhecido como a versão das licenças “legível por humanos”). O Resumo Explicativo é uma referência útil tanto para os licenciadores como para os licenciados, resume e expressa alguns dos termos e condições mais importantes. Você pode considerar o Resumo Explicativo como uma interface amigável com o Texto Legal subjacente, embora o Resumo não seja, em si mesmo, uma licença e o seu conteúdo não forme parte do Texto Legal propriamente dito.

3. LEGÍVEL POR MÁQUINAS

A camada final das licenças reflete o fato de que o software, desde motores de pesquisa, passando pelos pacotes de produtividade no escritório, até à edição de música, desempenha um papel enorme na criação, cópia, descoberta e distribuição de trabalhos. Para que a Internet identifique facilmente quando um trabalho está disponível sob uma licença Creative Commons, disponibilizamos uma versão digital (“legível por máquinas”) da licença – um resumo dos direitos e obrigações, expresso num formato que as aplicações informáticas, motores de pesquisa e outros tipos de tecnologia, possam compreender. Para consegui-lo, desenvolvemos uma forma padronizada de descrever as licenças que pode ser lida e entendida por software, chamada Linguagem de Expressão de Direitos, da CC (CC REL)..

Fonte: Creative Commons (2024)

APÊNDICE B - AS LICENÇAS DO CREATIVE COMMONS



AS LICENÇAS DO CREATIVE COMMONS

ATRIBUIÇÃO (CC BY)

Esta licença permite que outros distribuam, remixem, adaptem e criem a partir do seu trabalho, mesmo para fins comerciais, desde que lhe atribuam o devido crédito pela criação original. É a licença mais flexível de todas as licenças disponíveis. É recomendada para maximizar a disseminação e uso dos materiais licenciados.

ATRIBUIÇÃO-COMPARTILHA IGUAL (CC BY-SA)

Esta licença permite que outros remixem, adaptem e criem a partir do seu trabalho, mesmo para fins comerciais desde que lhe atribuam o devido crédito e que licenciem as novas criações sob termos idênticos. Esta licença costuma ser comparada com as licenças de software livre e de código aberto "copyleft". Todos os trabalhos novos baseados no seu terão a mesma licença, portanto quaisquer trabalhos derivados também permitirão o uso comercial. Esta é a licença usada pela Wikipédia e é recomendada para materiais que seriam beneficiados com a incorporação de conteúdos da Wikipédia e de outros projetos com licenciamento semelhante.

ATRIBUIÇÃO-SEMDERIVAÇÕES (CC BY-ND)

Esta licença permite a redistribuição, comercial e não comercial, desde que o trabalho seja distribuído inalterado e no seu todo, com crédito atribuído a você.

ATRIBUIÇÃO-SEMDERIVAÇÕES (CC BY-ND)

Esta licença permite que outros remixem, adaptem e criem a partir do seu trabalho para fins não comerciais, e embora os novos trabalhos tenham de lhe atribuir o devido crédito e não possam ser usados para fins comerciais, os usuários não têm de licenciar esses trabalhos derivados sob os mesmos termos.

ATRIBUIÇÃO-NÃOCOMERCIAL-COMPARTILHAIGUAL (CC BY-NC-SA)

Esta licença permite que outros remixem, adaptem e criem a partir do seu trabalho para fins não comerciais, desde que atribuam a você o devido crédito e que licenciem as novas criações sob termos idênticos.

ATRIBUIÇÃO-SEMDERIVAÇÕES-SEMDERIVADOS (CC BY-NC-ND)

Esta é a mais restritiva das nossas seis licenças principais, só permitindo que outros façam download dos seus trabalhos e os compartilhem desde que atribuam crédito a você, mas sem que possam alterá-los de nenhuma forma ou utilizá-los para fins comerciais.

Fonte: Creative Commons (2024)

SOBRE OS AUTORES E AS AUTORAS

ADRIANA ALVES RODRIGUES

Pós-Doutora e Doutora em Ciência da Informação pela Universidade Federal da Paraíba (UFPB). Mestre em Comunicação e Cultura Contemporâneas (linha Cibercultura) pela Universidade Federal da Bahia - UFBA, onde integrou o Grupo de Pesquisa em Jornalismo Online (GJol). Especialista em Jornalismo Contemporâneo pela Unijorge. Graduada em Comunicação Social - Jornalismo pela Universidade Estadual da Paraíba - UEPB; Vencedora dos Prêmios: Melhor Tese de Doutorado em Ciência da Informação (UFPB), em 2019, Prêmio PRPG de Teses da UFPB; Pesquisadora dos Grupos de Pesquisa em Jornalismo e Mobilidade (MOBJOR/UEPB). Leciona e Pesquisa nas áreas da Comunicação e Ciência da Informação com ênfase em Cibercultura, Jornalismo Digital, Jornalismo de Dados, Narrativas Multimídia, Visualização de Dados, Big Data, Dados, Informação e Tecnologia.

GUILHERME ATAÍDE DIAS

Pós-Doutor em Ciência da Informação pela Universidade Estadual Paulista (UNESP); Doutor em Ciência da Informação, com ênfase em Ciências da Comunicação, pela Universidade de São Paulo (USP); Mestrado em Gestão Organizacional pela Central Connecticut State University, USA (CCSU) (1995); Bacharelado em Ciência da Computação pela Universidade Federal da Paraíba (UFPB), Campus II (1990); Bacharelado em Direito pelo Centro Universitário de João Pessoa (UNIPÊ); Atualmente, ocupa a posição de Professor Titular na Universidade Federal da Paraíba, onde atua no Departamento de Ciência da Informação. No âmbito da pós-graduação, contribui para o Programa de Pós-Graduação em Ciência da Informação (PPGCI) e para o Programa de Pós-Graduação em Gestão nas Organizações Aprendentes (PPGOA), ambos na UFPB.

RAFAEL LUÍS GALDINI RAIMUNDO

Pós-Doutor pela Universidade de São Paulo (USP) e Doutor em Ciências (Ecologia) pela Universidade de São Paulo (USP); Bacharel em Ciências Biológicas e Mestre em Ecologia pela Universidade Estadual de Campinas (UNICAMP); Professor Adjunto no Departamento de Engenharia e Meio Ambiente da Universidade Federal da Paraíba (UFPB) e coordenador do Programa de Pós-Graduação em Ecologia e Monitoramento Ambiental (PPGEMA/UFPB). É credenciado no Programa de Pós-Graduação em Ciências Biológicas - PPGCB/UFPB. Coordenador do projeto DATAPB (<http://datapb.ccae.ufpb.br>) e do Laboratório Misto Internacional sobre Sustentabilidade IDEAL (<http://www.ideal.ufpb.br>). Atua nas áreas de Ecologia Evolutiva e Ciência da Sustentabilidade, desenvolvendo abordagens que buscam integrar modelos matemáticos e dados empíricos para aprofundar as bases teóricas de aplicações do conhecimento ecológico à conservação, desenvolvimento humano e resiliência econômica.

NOTA À EDIÇÃO

Esta obra que você, leitor, tem em mãos foi contemplada pelo Edital PRPG/UFPB Nº 01/2024, financiado pelo Programa de Apoio à Produção Científica - PRÓ-PUBLICAÇÃO DE LIVROS da Pró-Reitoria de Pós-Graduação, uma parceria entre a Editora UFPB e a PRPG. Ela representa o esforço de diversos pesquisadores e pesquisadoras, docentes, servidores técnico-administrativos, alunos e alunas desta instituição para divulgar o conhecimento científico produzido pela Universidade Federal da Paraíba.

O edital possibilitou a publicação de 13 livros em formato eletrônico sobre as mais variadas temáticas, reunindo pesquisadores ligados a dez departamentos, vinculados a sete diferentes centros de ensino e a dois campi da UFPB.

Das ciências das religiões às ciências da saúde, passando pelos estudos literários e sociais, apresentando reflexões sobre o fazer científico e os desafios educacionais, os títulos contemplados este ano apresentam um retrato - parcial e incompleto, visto que não contempla toda a pesquisa realizada na UFPB, mas ainda assim bastante significativo - da contribuição que nossa Instituição oferece à sociedade brasileira no intuito de avançar o fazer científico e ajudar no desenvolvimento do País.

Evandro Leite de Souza
Pró-Reitor de Pós-Graduação

Geysa Flávia Câmara de Lima Nascimento
Diretora Geral da Editora UFPB



GESTÃO DE DADOS DE PESQUISA

Título Ampliando a geração de conhecimentos científicos no contexto da ciência aberta

Autores(as) Adriana Alves Rodrigues
Guilherme Ataíde Dias
Rafael Luís Galdini Raimundo

Projeto gráfico e Capa Wellington Costa

Formato e-book (PDF – 16x22 cm)

Tipografia Myriad Pro

Número de páginas 95

