



**UNIVERSIDADE FEDERAL DA BAHIA
INSTITUTO DE CIÊNCIA DA INFORMAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO
MESTRADO EM CIÊNCIA DA INFORMAÇÃO**

UILLIS DE ASSIS SANTOS

**GESTÃO DE DADOS CIENTÍFICOS NAS BIBLIOTECAS DAS UNIVERSIDADES
FEDERAIS DO NORDESTE BRASILEIRO: ANÁLISE À LUZ DA CIÊNCIA
ABERTA**

Salvador
2025

UILLIS DE ASSIS SANTOS

**GESTÃO DE DADOS CIENTÍFICOS NAS BIBLIOTECAS DAS UNIVERSIDADES
FEDERAIS DO NORDESTE BRASILEIRO: ANÁLISE À LUZ DA CIÊNCIA
ABERTA**

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Informação da Universidade Federal da Bahia, como requisito para obtenção do título de Mestre em Ciência da Informação.

Linha de pesquisa 1: Políticas e Tecnologias da Informação.

Orientadora: Prof.^a Dr^a. Kátia de Oliveira Rodrigues.

Salvador
2025

Sistema Universitário de Bibliotecas - UFBA

Santos, Uillis de Assis

Gestão de dados científicos nas bibliotecas das universidades federais do Nordeste brasileiro: análise à luz da ciência aberta / Uillis de Assis Santos. - Salvador, 2025.

192 p.: il.

Orientador: Prof.^a Dr^a. Kátia de Oliveira Rodrigues

Dissertação (Mestre em Ciência da Informação) Universidade Federal da Bahia, Instituto de Ciência da Informação, Programa de Pós-Graduação em Ciência da Informação, 2025.

1. Gestão de dados científicos. 2. Bibliotecas universitárias – Brasil, Nordeste. 3. Serviços e produtos. 4. Ciência aberta. I. Rodrigues, Kátia de Oliveira. II. Título.

CDD – 612.115

Elaborada por Evandro Ramos dos Santos CRB-5/1205

UILLIS DE ASSIS SANTOS

Gestão de Dados Científicos nas Bibliotecas das Universidades Federais do Nordeste Brasileiro: análise à luz da ciência aberta

Dissertação apresentada ao Curso de Mestrado em Ciência da Informação do Programa de Pós-Graduação em Ciência da Informação (PPGCI) da Universidade Federal da Bahia (UFBA), como requisito para obtenção de grau de Mestre em Ciência da Informação.

Aprovada em: 21/07/25

Banca Examinadora



Profª. Dra. Kátia de Oliveira Rodrigues - Orientadora - UFBA



Documento assinado digitalmente
FABIANO COUTO CORRÉA DA SILVA
Data: 18/09/2025 22:25:45-0300
Verifique em <https://validar.iti.gov.br>

Prof. Dr. Fabiano Couto Corrêa da Silva - Membro Externo Titular - UFRS



Profª. Dra. Nídia Maria Lienert Lubisco - Membro Interno Titular – UFBA

Dedico esta dissertação à minha mãe,
Vera Lúcia de Assis Santos (*in memoriam*), pois sem
os seus ensinamentos essa trajetória não seria possível.

À minha esposa, Diana Paula, e ao meu filho, Saulo Akin.

AGRADECIMENTOS

Agradeço, primeiramente, à força superior que rege todas as coisas do Universo e que me sustenta desde sempre.

À minha esposa, Diana Paula, e ao meu filho, Saulo Akin, pelo amor, carinho e paciência durante os momentos mais difíceis dessa trajetória, sem vocês, eu não conseguiria chegar até aqui.

À minha orientadora, Prof.^a Kátia de Oliveira Rodrigues, pela confiança, respeito, condução humana e palavras de lucidez em meio ao turbilhão do meu percurso acadêmico.

Aos meus irmãos, Ulisses, Joseane, Josenilza, e aos meus sobrinhos, Lucas e Luigi.

Ao meu amigo Daniel Harry, por ter me incentivado na escolha pela profissão de bibliotecário.

Aos amigos do SIBIUFBA, Kleber Carvalho, Fernanda Xavier, Patrícia Barroso, Nara Silva e Érica Santana, pelo apoio e sugestões.

À Prof.^a Ivana Lins, pelo incentivo, ensinamentos e apoio para o desenvolvimento da Ciência Aberta no Sistema de Bibliotecas da UFBA.

À amiga Juliana Lima, da Universidade Federal do Ceará (UFC), por ter me inspirado com a sua dissertação a querer desbravar o mundo dos dados científicos.

Ao amigo Rogério Marques, pelos conselhos e palavras de incentivo.

Aos amigos da Rede Nordeste de Repositórios, Cristiane Estevão (UFAL), Fábio Gomes (UFRB), Silvana Araújo (UNILAB), Juliana Lima (UFC), Luhilda Silveira (UFMA), Rogério Marques e Ana Lopes (UFPB), Johnny Rodrigues (UFCG), Susimery Vila Nova (UFPE), Vânia Vida (UFRPE), Gisela Lima (UFPI), Clediane Marques (UFRN) e Mario Gaudêncio (UFERSA), pela colaboração na coleta dos dados.

E, por fim, a todos aqueles que contribuem para o avanço da Ciência Aberta.

SANTOS, Uillis de Assis. **Gestão de dados científicos nas bibliotecas das universidades federais do Nordeste brasileiro: análise à luz da Ciência Aberta.** 2025. Orientadora: Kátia de Oliveira Rodrigues. 190 p. il. Dissertação (Mestrado em Ciência da Informação) – Instituto de Ciência da Informação, Universidade Federal da Bahia, Salvador.

RESUMO

A dinâmica do contexto científico centrado em dados iniciado no século XX promoveu novas perspectivas para o ecossistema científico, com o reconhecimento dos dados como insumos para a retroalimentação da pesquisa na ciência. Diante dessa nova configuração, emergiram exigências para o compartilhamento dos conjuntos de dados e a construção de Planos de Gestão de Dados como requisitos dos periódicos científicos e das agências de fomento à pesquisa. Sob esse prisma, as bibliotecas universitárias, como mediadoras do conhecimento, passaram a desempenhar um papel crucial na formulação de serviços e produtos para esse gerenciamento no ambiente das Instituições de Ensino Superior. Foi nesse contexto de mudanças no âmbito da produção e da comunicação científica que a presente pesquisa estabeleceu a seguinte **questão de partida**: - Qual o panorama de promoção de serviços e produtos para o gerenciamento de dados científicos evidenciados nos sistemas de bibliotecas das universidades federais do Nordeste brasileiro? E como **objetivo geral** propôs: diagnosticar como os sistemas de bibliotecas das universidades federais do Nordeste brasileiro estão promovendo ou se preparando para promover a gestão dos dados científicos das suas instituições. Para alcançá-lo, elegeram-se os seguintes **objetivos específicos**: a) mapear, a partir da literatura científica, quais são os serviços e produtos pertinentes para o gerenciamento de dados científicos; b) averiguar nos *sites* de bibliotecas de instituições de ensino superior estrangeiras a existência de serviços e produtos destinados ao gerenciamento de dados científicos, contribuindo com informações complementares para elaboração do diagnóstico; c) identificar em qual estágio se encontram os serviços e produtos para gestão de dados científicos nos sistemas de bibliotecas das universidades do Nordeste brasileiro. Quanto ao **método**, procedeu-se ao levantamento bibliográfico para construção do referencial teórico; quanto aos objetivos, trata-se de uma pesquisa descritiva, com abordagem qualitativa e quantitativa; quanto aos procedimentos, adotou-se a pesquisa documental; quanto à técnica, optou-se pela observação sistemática e pela aplicação de questionário para as direções e coordenações desses sistemas de bibliotecas. A amostra foi composta por 14 sistemas de bibliotecas de universidades federais do Nordeste brasileiro, que revelaram baixa oferta de serviços e produtos para o gerenciamento de dados científicos, alta porcentagem na ausência de planejamento para que os sistemas de bibliotecas investigadas disponibilizem futuramente os serviços e produtos que ainda não são ofertados e desconhecimento, por parte de algumas coordenações e diretorias dos sistemas de bibliotecas, sobre gestão de dados científicos. O presente cenário remete para a necessidade da adoção de diretrizes e práticas alinhadas aos princípios da Ciência Aberta, focada na gestão dos dados científicos, conforme experiências coletadas em 3 universidades estrangeiras.

Palavras-chave: gestão de dados científicos - serviço e produto; gestão de dados científicos - sistema de biblioteca universitária; Ciência Aberta.

SANTOS, Uillis de Assis. **Research data management in the library systems of federal universities in the Brazilian Northeast: analysis in light of Open Science.** 2025. Adviser: Kátia de Oliveira Rodrigues. 190 p. il. Thesis (Masters in Information Science) – Institute of Information Science, Federal University of Bahia, Salvador, 2025.

ABSTRACT

The dynamics of the data-centric scientific context, which began in the 20th century, fostered new perspectives for the scientific ecosystem, with the recognition of data as input for research feedback in science. Faced with this new configuration, demands emerged for the sharing of datasets and the creation of Data Management Plans as requirements from scientific Journals and research funding agencies. From this perspective, university libraries, as knowledge mediators, began to play a crucial role in formulating services and products for the research data management within Higher education Institutions. It is within this context of significant changes in scientific production and communication that the present research sought to investigate the landscape of promoting services and products for research data management evident in the library systems of federal universities in Brazilian Northeast. The General Objective proposed: to diagnose how library systems of federal universities in the Brazilian Northeast are promoting or preparing to promote the research data management from their institutions. To achieve this, the following specific objectives were chosen: a) to map, from the scientific literature, which services and products exits for research data management; b) to investigate the website of libraries of foreign higher education institutions for the existence of services and products aimed at research data management, contributing complementary information for the preparation of the diagnosis; c) to identify the stage of services and products for research data management in the library systems of universities in the Brazilian Northeast. The methodology is descriptive with a qualitative and quantitative approach, and in terms of technical procedures it is classified as a literature review, documentary research, and survey. The sample consisted of fourteen library systems of federal universities in Brazilian Northeast, with a questionnaire applied to the directors and coordinators of these library systems. The research revealed a low offer of services and products for research data management, high percentage absence of planning for the investigated library systems to offer currently unavailable services and products in the future, and a lack of knowledge among some coordinators and directors of library systems regarding research data management. The current scenario highlights the need to adopt guidelines and practices aligned with the principles of Open Science focused on research data management.

Keywords: research data management – service and product; research data management - university library systems; Open Science.

LISTA DE FIGURAS

Figura 1	Ecossistema científico da <i>Center for Open Science</i>	24
Figura 2	Ecossistema da coleta de metadados do <i>Oasisbr</i>	30
Figura 3	Taxonomia da Ciência Aberta <i>FOSTER</i>	42
Figura 4	Status dos Planos de Dados Abertos.....	81
Figura 5	Pergunta da primeira versão do questionário.....	113
Figura 6	Pergunta da versão final do questionário.....	113
Figura 7	FAQ sobre a Política do <i>NIH</i>	116
Figura 8	Workshop dados governamentais na pesquisa científica.....	117
Figura 9	Boletim eletrônico.....	117
Figura 10	<i>Camp Data Management</i>	118
Figura 11	Evento sobre citação de dados.....	119
Figura 12	Evento sobre gestão de dados.....	119
Figura 13	Palestra sobre a Política da <i>NIH</i>	120
Figura 14	<i>Research Data Management Librarian Data Academy</i>	122
Figura 15	Explicando os Princípios FAIR.....	123
Figura 16	Aplicação dos Princípios FAIR.....	123
Figura 17	Workshop sobre os Princípios FAIR.....	124
Figura 18	Curso sobre o ciclo de vida dos dados.....	124
Figura 19	<i>Open Science Framework (Harvard University)</i>	125
Figura 20	<i>Open Science Framework (MIT)</i>	125
Figura 21	Ciclo de Vida dos Dados na área Biomédica.....	126
Figura 22	<i>DMPTool (Harvard University)</i>	127
Figura 23	<i>DMPTool (MIT)</i>	127
Figura 24	<i>DMPonline (University of Cambridge)</i>	127
Figura 25	Curso para bibliotecários sobre o PGD.....	128
Figura 26	Evento sobre dados sensíveis.....	128
Figura 27	FAQ sobre dados pessoais e sensíveis.....	129
Figura 28	Informações sobre dados pessoais e sensíveis.....	129
Figura 29	Mesa redonda sobre mineração de textos e dados.....	130
Figura 30	Metadados para encontrar, interpretar e compartilhar dados.....	131
Figura 31	Repositório de Dados Científico (<i>Apollo</i>).....	132

Figura 32	Repositório de Dados Científico (<i>Harvard Dataverse</i>)	132
Figura 33	Repositório de Dados Científico da <i>ICPSR</i>	133
Figura 34	Repositório de Dados Científico da <i>QDR</i>	133
Figura 35	Curso para encontrar repositórios de dados científicos.....	134
Figura 36	Workshop depositando no Repositório <i>Dryad</i>	134
Figura 37	Repositório de Dados de Pesquisa Antônio Martins Filho (REDAM)...	147

LISTA DE QUADROS

Quadro 1 - Mudança cultural da <i>Center for Open Science</i>	24
Quadro 2 - Serviços e produtos ofertados pela OKF	26
Quadro 3 - Princípios FAIR	27
Quadro 4 - Produtos, serviços e eventos criados pelo IBICT	31
Quadro 5 - Cursos em Ciência Aberta da Fiocruz	33
Quadro 6 - Movimentos do Ecossistema da Ciência Aberta	45
Quadro 7 - Declarações do Movimento de Acesso Aberto	53
Quadro 8 - 4 ° Plano de Ação Nacional para o Governo Aberto.....	82
Quadro 9 - Classificação dos dados	85
Quadro 10 - Política mandatória para a formulação do PGD.....	91
Quadro 11 - Classificação dos repositórios de dados científicos.....	96
Quadro 12 - Tratamento dos dados pessoais e sensíveis pela LGPD.....	98
Quadro 13 - Primeira fase da seleção da amostra da pesquisa	108
Quadro 14 - Segunda fase da seleção da amostra da pesquisa	109
Quadro 15 - Universidades estrangeiras selecionadas em ranking internacionais..	111
Quadro 16 - Relatos dos sistemas de bibliotecas	149
Quadro 17 - Bibliotecas Estrangeiras X Bibliotecas do Nordeste Brasileiro.....	152

LISTA DE GRÁFICOS

Gráfico 1	Política/Norma para gestão de dados científicos.....	136
Gráfico 2	Divulgação de informações sobre GDC.....	138
Gráfico 3	Palestras sobre GDC.....	139
Gráfico 4	Verba para capacitação de bibliotecários em GDC.....	140
Gráfico 5	Orientação sobre Princípios FAIR.....	141
Gráfico 6	Orientação sobre o ciclo de vida dos dados.....	142
Gráfico 7	Orientação para o preenchimento do Plano de Gestão de Dados.....	143
Gráfico 8	Orientação para anonimização dos dados.....	144
Gráfico 9	Orientação para encontrar dados úteis.....	145
Gráfico 10	Repositório de dados científico.....	146
Gráfico 11	Orientação para indicar repositório de dados científicos.....	149

SUMÁRIO

1 INTRODUÇÃO	12
2 FAZER CIENTÍFICO CENTRADO EM DADOS	16
3 CIÊNCIA ABERTA NO MUNDO E NO BRASIL	22
3.1 Ciência Aberta: conceitos, características e taxonomias	37
3.2 Movimento de Acesso Aberto <i>versus</i> oligopólio editorial	47
3.2.1 Declarações, manifestos e infraestruturas eletrônicas do Movimento de Acesso Aberto	52
4 A IMPORTÂNCIA DOS DADOS NA SOCIEDADE CONTEMPORÂNEA	57
4.1 Dados científicos	60
5 BIBLIOTECAS UNIVERSITÁRIAS NO SUPORTE À GESTÃO DE DADOS CIENTÍFICOS	65
5.1 Editoras científicas	71
5.2 Políticas para o compartilhamento de dados científicos	74
5.3 Classificação dos dados	84
5.4 Ciclo de vida dos dados	87
5.5 Plano de gestão de dados	90
5.6 Repositórios de dados científicos	95
6 PERCURSO METODOLÓGICO	105
6.1 Universo e amostra da pesquisa	106
6.2 Instrumentos e procedimentos da coleta de dados	110
6.3 Pré-teste	112
6.4 Análise dos dados	114
7 APRESENTAÇÃO E ANÁLISE DOS RESULTADOS	115
7.1 Serviços e produtos para GDC em bibliotecas estrangeiras	115

7.2 Identificação de produtos e serviços nos sistemas de bibliotecas das universidades federais do Nordeste brasileiro	135
8 CONSIDERAÇÕES FINAIS	151
REFERÊNCIAS	157
APÊNDICE A - QUESTIONÁRIO	179
APÊNDICE B – PLANO DE GESTÃO DE DADOS	183
APÊNDICE C – PROTOCOLO DE COLETA DE DADOS DAS BIBLIOTECAS ESTRANGEIRAS	187
APÊNDICE D – SERVIÇOS E PRODUTOS PARA GESTÃO DE DADOS CIENTÍFICOS (SPGDC)	188
ANEXO A – TAXONOMIA DA CIÊNCIA ABERTA CONFORME PESQUISADORES BRASILEIROS	191
APÊNDICE B – TAXONOMIA DA CIÊNCIA ABERTA	192

1 INTRODUÇÃO

A introdução dessa pesquisa, inevitavelmente, remete à discussão das mudanças ocorridas nas últimas décadas na produção e disseminação da informação no campo científico, utilizando como referência o uso intensivo dos dados. Indiscutivelmente, o advento do Quarto Paradigma¹ revolucionou as práticas científicas, tornando-se também um dos principais combustíveis para a promoção das transformações em diversos setores como economia, política, comunicação, agricultura e saúde.

No tocante ao setor científico, as Tecnologias de Informação e Comunicação (TIC) provocaram uma aceleração nas capacidades de produção, processamento e disseminação da informação. Nessa perspectiva, foi criado um mecanismo cíclico e contínuo, pautado na gestão da inovação e da mudança para a produção de mais informação e conhecimento (Castells, 1999). A sociedade experimentou revoluções na oferta de informação e, por consequência, uma maior interconectividade. Entretanto, essa revolução também acarretou a exclusão de diversos indivíduos em situação de fragilidade econômica ao acesso à informação, por conta dos mecanismos do capitalismo que estão baseados na desigual partilha daquilo que é produzido e comercializado na sociedade, gerando, assim, acúmulo de riquezas para uma pequena parte da população.

As influências do capitalismo também alcançaram os canais de disseminação da informação da comunidade científica. Nesse processo, o conhecimento científico foi privatizado pelos grandes oligopólios editoriais, por meio das ferramentas de retenção dos direitos autorais dos pesquisadores e das cobranças abusivas para permitir o acesso aos artigos científicos. Em contrapartida, a comunidade científica organizou-se para criar alternativas às práticas desse oligopólio, com a concepção do Movimento de Acesso Aberto e, posteriormente, com a estruturação da Ciência Aberta.

Durante toda essa efervescência de mudanças paradigmáticas, os dados foram ocupando um espaço de destaque, principalmente no contexto da ciência. O

¹ O Quarto Paradigma se caracteriza pela unificação das teorias de experimento e simulação, da captura de dados por instrumentos ou gerados por simuladores, processamento de dados por software, informação/conhecimento armazenado em computador e cientistas que analisam banco de dados/arquivos usando gerenciamento de dados e estatísticas (Hey; Tansley; Tolle, p. 18, 2009).

reconhecimento do potencial dos dados como insumos de retroalimentação das engrenagens do ecossistema científico acarretou a construção de políticas, métodos, técnicas e tecnologias específicas para o tratamento adequado dessa poderosa fonte de informação.

Indiscutivelmente, é nesse contexto de mudanças que as bibliotecas universitárias precisam incluir-se para desempenhar seu papel de fomentadoras da informação, independentemente do suporte em que essa informação se encontre. Todavia, para proceder com o tratamento e a criação de serviços dessa complexa fonte de informação, faz-se necessário um redimensionamento das habilidades dos profissionais em biblioteconomia.

A complexidade dos dados científicos advém de uma série de questões que ultrapassam as tradicionais fontes de informações padronizadas e estruturadas (livros, artigos, dissertações, teses e relatórios científicos), nas quais os bibliotecários foram preparados para atuar. Os dados possuem uma complexidade inerente, oriunda dos diferentes campos científicos que os criaram, além do uso de diferentes métodos e tecnologias no momento de sua concepção, o que confere uma heterogeneidade em seus formatos (Sales *et al.*, 2019). Para coletar, tratar, disponibilizar e preservar esta complexa fonte de informação, o contexto exige abordagens holísticas e novas *expertises* por parte dos bibliotecários.

Analizando o cenário descrito anteriormente, a presente pesquisa justifica-se pelas possíveis contribuições que poderá proporcionar ao campo da Ciência da Informação e da Ciência, como, por exemplo, a adaptação das tradicionais técnicas de tratamento da informação existentes na biblioteconomia para serem aplicadas ao tratamento dos dados e a inserção de temáticas ainda pouco trabalhadas na área.

O interesse do pesquisador pela temática dos dados científicos surgiu a partir do ano de 2018, enquanto gerenciava o Repositório Institucional de Publicações Científicas da Universidade Federal da Bahia (RI/UFBA).

As solicitações dos pesquisadores da UFBA, enviadas ao e-mail do Repositório, com dúvidas sobre o depósito dos conjuntos de dados de suas pesquisas e a construção do Plano de Gestão de Dados (PGD), foram o ponto de partida para adentrar nos estudos sobre os dados científicos e para compreender como as bibliotecas universitárias poderiam se inserir no contexto da oferta de serviços e produtos voltados para o gerenciamento desses dados científicos. Posteriormente,

com o objetivo de aprofundar ainda mais o conhecimento sobre a temática e desenvolver pesquisa que abordasse questões relacionadas à gestão de dados científicos, optou-se pelo ingresso no Programa de Pós-Graduação em Ciência da Informação (PPGCI) pela UFBA, como um caminho possível para alcançar êxito nesta inquietação.

Nessa perspectiva, desenvolveu-se um estudo com a seguinte **questão de pesquisa**: qual o panorama de promoção de serviços e produtos para o gerenciamento de dados científicos evidenciados nos sistemas de bibliotecas das universidades federais do Nordeste brasileiro?

Nesse desdobramento, constituiu-se como **objetivo geral** da pesquisa: diagnosticar como os sistemas de bibliotecas das universidades federais do Nordeste brasileiro estão promovendo ou se preparando para promover serviços e produtos para gestão dos dados científicos das suas instituições. Para alcançar o objetivo geral, foram delineados os seguintes **objetivos específicos**: a) mapear, a partir da literatura científica, quais os serviços e produtos para o gerenciamento de dados científicos; b) averiguar nos *sites* de bibliotecas de instituições de ensino superior estrangeiras a existência de serviços e produtos destinados ao gerenciamento de dados científicos, contribuindo com informações complementares para elaboração do diagnóstico; c) identificar em qual estágio se encontram os produtos e serviços para gestão de dados científicos nos sistemas de bibliotecas das universidades do Nordeste brasileiro.

Diante desses pressupostos, a pesquisa foi classificada em seu nível como descritiva, a abordagem é qualitativa e quantitativa e, quanto aos procedimentos técnicos, é classificada como uma pesquisa documental, e de levantamento. Para alcançar os elementos descritos anteriormente, a dissertação foi estruturada em oito seções.

A primeira seção constitui-se da introdução: contextualização da pesquisa, a justificativa, o problema e os objetivos.

A segunda seção apresenta as mudanças paradigmáticas no fazer científico durante os séculos XX e XXI, centrada em dados, e em seguida discute os movimentos surgidos em diversas partes do mundo no contexto da produção e tratamento dos dados científicos.

A terceira seção demonstra a consolidação da Ciência Aberta como movimento central que engloba diversas mudanças paradigmáticas do ecossistema científico,

aponta algumas instituições envolvidas com o desenvolvimento da Ciência Aberta no mundo e no Brasil. Expõe os conceitos, as taxonomias, discussões teóricas e as influências da Ciência Aberta no ecossistema da produção e disseminação científica. Apresenta também os movimentos componentes da Ciência Aberta e discorre com um maior aprofundamento sobre o Movimento de Acesso Aberto no que tange ao seu surgimento através de declarações, manifestos e infraestruturas eletrônicas.

A quarta seção expõe a importância que os dados alcançaram em diversos setores da sociedade. Apresenta os conceitos acerca dos dados científicos, além de discutir a produção e tratamento dos dados no âmbito dos grandes e pequenos projetos científicos (Cauda Longa da Ciência).

A quinta seção aborda a importância do gerenciamento adequado dos dados científicos para o desenvolvimento da ciência. Apresenta a necessidade da atuação das bibliotecas universitárias como protagonista do processo de planejamento e criação de Serviços e Produtos para o Gerenciamento de Dados Científicos (SPGDC). Demonstra bibliotecas universitárias estrangeiras que já estão consolidadas na prestação de suporte ao gerenciamento de dados científicos, expõe elementos para a promoção do gerenciamento dos dados científicos extraídos da literatura científica e discute sobre algumas informações do arcabouço da Ciência Aberta, focadas em dados científicos que são fundamentais para as bibliotecas promoverem os SPGDC.

Na sexta seção, é apresentada como a pesquisa está classificada, a metodologia, as etapas para seleção da amostra e os critérios para a construção do instrumento de coleta de dados.

Na sétima seção, é realizada a exposição das análises dos resultados revelados pela pesquisa.

Na oitava seção, são apresentadas as considerações finais, com a configuração do cenário dos sistemas de bibliotecas das universidades do Nordeste brasileiro em relação à oferta dos serviços e produtos para gestão de dados científicos, os benefícios que podem ser gerados com a oferta desses serviços e produtos, limitações da pesquisa e recomendações para pesquisas futuras.

2 FAZER CIENTÍFICO CENTRADO EM DADOS

A sociedade, no século XX, vivenciou profundas transformações econômicas, sociais e culturais com a explosão informacional pós-Segunda Guerra Mundial e, posteriormente, com a concepção de tecnologias sofisticadas em todos os setores (Castells, 1999). O uso de *software* e *hardware* de alta performance propiciou um aumento exponencial de dados no campo da ciência durante o século XX e, consequentemente, provocou mudanças significativas no fazer científico. Conceitos, teorias e procedimentos consolidados, outrora, passaram a ser aprimorados por perspectivas advindas de novas observações e experimentos baseados no uso dos dados.

É nesse contexto de remodelagem de práticas e conceitos que ocorrem o desenvolvimento da ciência e a construção de novos paradigmas nos diversos campos científicos. O paradigma, segundo Kuhn (1975, p. 13), pode ser descrito como: “[...] as realizações científicas universalmente reconhecidas que, durante algum tempo, fornecem problemas e soluções modelares para uma comunidade de praticantes de uma ciência”. Nesse sentido, entende-se que o estabelecimento de paradigmas é fundamental para a organização da ciência, porém, sem a sua ruptura, é impossível alcançar novas possibilidades. A visão de Oliveira e Silva (2016, p. 9), corrobora a reflexão anterior quando exemplifica que “as discussões acerca dos paradigmas destacam que quando há uma grande mudança no *statu quo* e no *modus operandi* do fazer científico ocorre uma transição ou rompimento com práticas anteriores”.

No entendimento de Gray, o fazer científico já vivenciou três paradigmas, denominados Ciência Empírica, Ciência Teórica, Ciência Computacional, e atualmente vivencia-se o quarto paradigma. De acordo com os autores, a Ciência Empírica remonta aos primórdios da ciência, e seu fazer baseia-se nas experiências sensoriais e observações; a Ciência Teórica é representada pelas leis, modelos e generalizações; a Ciência Computacional é representada pelo uso de computadores para simular complexas teorias. Quanto ao quarto paradigma, a predominância é o uso de tecnologias sofisticadas para a geração e o tratamento de uma quantidade de dados jamais vista, assim como o estabelecimento de um intrincado processo

colaborativo de pesquisadores dos diferentes campos científicos, por meio de uma rede eletrônica (Hey; Tansley; Tolle, 2009).

O Quarto Paradigma ou Ciência Intensiva em Dados seria o fio condutor que interliga os paradigmas anteriores, fornecendo, a partir de sua estrutura, um ciclo de apoio mútuo entre os paradigmas para que os cientistas alcancem seus objetivos no percurso de sua pesquisa (Lynch, 2009). Logo, o Quarto Paradigma se configura para promover novas possibilidades ao fazer científico. Oliveira e Silva (2016, p. 1) relatam que a nova forma de desenvolver ciência “[...] fomenta uma metodologia científica orientada aos dados de pesquisa. Assim, o dado é o elemento central e o seu recurso principal”.

É oportuno apresentar que os desafios inerentes ao fazer científico centrado em dados não estão presentes apenas nas dificuldades enfrentadas na gestão dos dados gerados pelos projetos científicos, mas também no campo conceitual e operacional advindo das multiplicidades de termos e ações dessa prática científica. Evidenciam-se os desafios nas diversas nomenclaturas que variam a depender do país ou campo científico, assim como das particularidades de cada uma delas: *e-Science*; *data-drivenscience*; *data-intensivecomputing*; *data-intensive science*; *e-Research*; *Cyberiscience*; *Cyberinfrastructure*; *data-mining* e *Open Science* (Gold, 2007; Wouters; Beaulieu, 2007; Hey; Tansley; Jankowski, 2009; Lynch, 2009; Tolle, 2009; César Júnior, 2011; Whitmire, 2013; *Organisation for Economic Co-Operation and Development*, 2015).

Nesse sentido, o fazer científico centrado em dados é um fenômeno que ainda está em processo de construção de seus conceitos, terminologias e práticas, como foi proferido por César Júnior (2011, p. 7):

Uma nova área científica nasceu nos últimos dez anos. Como ocorre frequentemente na história da ciência, ela nasceu em diferentes lugares, com diferentes nomes, mas da necessidade comum de se enfrentar o dilúvio de dados.

Apesar da relevância de todas as iniciativas listadas acima, destacaremos a *e-Science*, *Cyberinfrastructure*, *e-Research* e *Open Science*. As três primeiras representam uma mudança na forma de fazer ciência, sendo que as duas primeiras surgem de projetos governamentais, evidenciando a importância do financiamento público para o desenvolvimento da ciência (Jankowski, 2007). Em relação à *Open*

Science, traduzida no Brasil como Ciência Aberta, chamam a atenção as suas contribuições revolucionárias na construção do fazer científico centrado em dados e por representar um dos pontos centrais desta pesquisa. Diante dessa diversidade terminológica, faz-se necessária uma breve apresentação sobre cada termo, com a exceção da Ciência Aberta que será abordada na seção 3.

O termo *e-Science* surgiu em 1999 e foi desenvolvido por Jonh Taylor, então Diretor-geral de Conselhos de Pesquisa do Escritório de Ciência e Tecnologia do Reino Unido (Jankowski, 2009; Marcum; George, 2010). A proposta almejava definir as novas práticas surgidas em diversas disciplinas, que evidenciaram um processo colaborativo multidisciplinar e o gerenciamento adequado dos dados como ponto central para o desenvolvimento da ciência no Reino Unido.

No site da *National e-Science Centre* do Reino Unido, pode-se encontrar a definição do termo elaborada por Jonh Taylor no ano de 1999, um vislumbre do que a *e-Science* poderia se tornar: “No futuro, *e-Science* se referirá à ciência em grande escala que será cada vez mais realizada por meio de colaborações globais distribuídas habilitadas pela internet”.

A definição de Jonh Taylor encontra consonância com a visão das particularidades em torno da *e-Science* de Marcum e George (2010, p. 4), onze anos depois.

[...] ciência computacional intensa que emprega tecnologia de computação em grande escala para analisar coleções de dados muito grandes. Ela pode ser ainda caracterizada por colaboração distribuída em larga escala através da internet combinada com tecnologia de suportes essenciais, como computação de alto desempenho e grande gerenciamento de informações em escala.

No ano de 2004, o governo do Reino Unido fundou o Centro Nacional para Ciências Sociais Eletrônicas, com o objetivo de promover a *e-Science* também no campo das Ciências Sociais por meio de 11 projetos-pilotos, já que o desenvolvimento da *e-Science*, inicialmente, foi bastante difundido no campo das Ciências Naturais (Jankowski, 2009). Os investimentos do governo britânico, por meio de capacitações, projetos e infraestrutura eletrônica foram os principais responsáveis pelo desenvolvimento da *e-Science* no Reino Unido.

Nos Estados Unidos da América (EUA), o fazer científico centrado em dados ficou conhecido como *Cyberinfrastructure* pelo Relatório *Atkins*, que *foi* encomendado pela *National Science Foundation (NSF)*. O relatório foi elaborado por diversos

cientistas que elaboraram recomendações para que a *National Science Foundation* pudesse liderar uma “[...] revolução na ciência e na engenharia por meio do desenvolvimento coordenado e do uso expansivo da infraestrutura cibernética” (Atkins et al., 2003, p. 32). O relatório apontava que a revolução das tradicionais práticas do fazer científico ocorreria por meio da computação digital, dados, informações e redes (Atkins et al., 2003). Nessa perspectiva, a necessidade de investimentos públicos era percebida como um fator fundamental para o desenvolvimento da pesquisa científica nos EUA.

Por meio do Relatório *Atkins*, a *Cyberinfrastructure* é definida como uma “infraestrutura baseada em computadores distribuídos, tecnologia de informação e comunicação” (Atkins et al., 2003, p. 5). Uma infraestrutura cibernética responsável pela união de elementos como: computação em grade, capacidade de armazenamento e preservação de grandes quantidades de dados, *software*, serviços, informações, conhecimentos, equipes multidisciplinares e instituições.

O Relatório *Atkins* foi fundamental para a adoção de estratégias e investimentos para a concepção de uma infraestrutura que possibilitou aos EUA inserirem-se nesse novo contexto baseado em dados, ao mesmo tempo que colocou o país como referência na gestão de dados científicos.

A *e-Research*, segundo Lucas e Willinsky (2009), surge inspirada nas recomendações do Relatório *Atkins* e da revolução das práticas científicas ocorrida nos EUA. O termo passou a ser bastante utilizado pelas bibliotecas universitárias dos EUA no início do século XXI. A *Association of Research Libraries (ARL)* definiu *e-Research* como “formas de pesquisa computacionalmente intensiva em grande escala e em rede colaborativa [...]” (Whitmire, 2013, p. 70). Nos Países Baixos, o termo *e-Research* foi utilizado pela primeira vez por Wonters e Bealieu, por meio das iniciativas da *Royal Netherlands Academy of Arts and Sciences (KNAW)* e a *The Virtual Knowledge Studio for the Humanities and Social Sciences (VKS)* (Wonters; Bealieu, 2007).

Wonters e Bealieu (2007) propuseram o uso do novo termo com a justificativa de que a *e-Science*, utilizada amplamente no Reino Unido, não contemplava as peculiaridades das áreas das Ciências Humanas e Sociais, isso porque, de acordo com os pesquisadores, a *e-Science* estava fortemente relacionada à estrutura computacional e com orientação para a gestão de dados com abordagem quantitativa.

De acordo com Wonters e Bealieu (2007), a *e-Research* propõe maior integração multidisciplinar para a geração de conhecimento, ao mesmo tempo que respeita e reconhece a existência das múltiplas formas de fazer ciência com o uso das tecnologias. Os processos colaborativos perpassam desde o desenvolvimento, execução e financiamento de um projeto, os quais somente são possíveis com o “resultado de um complexo processo de interação entre as dimensões social, política, econômica, tecnológica e epistêmica do processo de pesquisa” (Wonters; Bealieu, p. 585, 2007).

Em contrapartida ao pensamento de Wonters e Bealieu, a bibliotecária Amanda L. Whitmire, em seu artigo *Thoughts on “eResearch”: a Scientist’s perspective*, direcionou várias críticas ao uso do termo *e-Research* pela comunidade bibliotecária dos Estados Unidos. A bibliotecária descreveu que o termo era totalmente desconhecido para a grande maioria dos pesquisadores. Assim, na perspectiva de Whitmire (2013, p. 69), a *e-Research*, quando foi utilizada para representar os serviços prestados pela comunidade bibliotecária no apoio à gestão de dados, acabou tornando-se um termo “[...] inconsistente e confuso, especialmente para não bibliotecários”, além de possuir similaridades conceituais com a *e-Science* e o fato de causar ainda mais desentendimento do termo. A visão de Whitmire é um exemplo das divergências interpretativas e conceituais acerca das transformações em decurso das práticas científicas.

Os conceitos apresentados anteriormente possuem a intenção de promover a contextualização do fazer científico centrado em dados na contemporaneidade e estabelecer uma perspectiva dos diferentes pontos de vista inseridos nesse processo. A multiplicidade interpretativa e conceitual de uma determinada temática é uma característica inerente ao universo da ciência, e suas consequências são igualmente transformadoras para o ser humano e o seu contexto social.

Na perspectiva de alguns teóricos, evidentemente, a ciência pode não estar vivenciando um novo paradigma. Segundo Oliveira e Silva (2016, p. 10), o atual contexto poderia ser “[...] uma confluência e aprimoramento de teorias, métodos, modelos, práticas e funcionalidades diferentes e revalidadas [com o apoio de novas tecnologias]”. Contudo, as teorias desenvolvidas por Kuhn (1975, p. 69), acerca do “paradigma” e das “regras” que regem uma determinada área do conhecimento,

revelam que “a falta de uma interpretação padronizada ou de uma redução a regras que goze de unanimidade não impede que um paradigma oriente a pesquisa”.

Indubitavelmente, o campo científico vivencia profundas transformações em suas práticas, como: a adoção de tecnologias de códigos abertos não proprietárias, disponibilização em acesso aberto da literatura científica, maior transparência nos processos metodológicos e avaliações científicas, construção de projetos científicos com equipes interdisciplinares, participação de pessoas de fora do contexto científico na construção da ciência, compartilhamento de *pré-prints* e a disponibilização dos dados científicos. Essas características são evidências das mudanças em curso na produção e comunicação científica.

3 CIÊNCIA ABERTA NO MUNDO E NO BRASIL

A Ciência Aberta passou a ser um termo frequentemente utilizado no início do século XXI pela comunidade científica para definir mudanças paradigmáticas ocorridas na ciência em seu fazer e comunicação, principalmente nas recomendações das práticas ancoradas nos movimentos de *Open Access* (Acesso Aberto) e *Open Source* (Código Aberto).

Em relação às outras iniciativas descritas na seção anterior, a Ciência Aberta é um termo mais recente. O termo incorpora, amplia e apresenta novas características aos conceitos já utilizados para denominar a nova forma de se desenvolver e comunicar a ciência em diversos campos científicos. Na visão de Quintanilha (2015, p. 14), “o percurso não foi imediato. De conceitos iniciais, como *ciber-ciência*, *ciberinfraestrutura*, *e-science*, *e-research*, o trajeto feito levou-nos à designação hoje conhecida como Ciência Aberta [...].” Naturalmente, essas construções conceituais refletem as necessidades paradigmáticas da ciência pela busca de novos caminhos para o fazer científico.

Os termos utilizados ao redor do mundo para denominar o fazer científico centrado em dados, assim como na Ciência Aberta, possuem pontos em comum que permeiam seus objetivos e desafios, como, por exemplo: o uso de tecnologias sofisticadas (*hardware* e *software*) para geração e manipulação de uma grande quantidade de dados; a construção de um ambiente de rede para propiciar a colaboração de pesquisadores em diferentes partes do mundo e o uso de infraestruturas eletrônicas para a distribuição das descobertas científicas. Mas, então, o que difere a Ciência Aberta das outras terminologias?

A principal característica é personificada pelo foco na abertura de todos os processos que envolvem o ecossistema científico. As recomendações abrangem não somente o acesso mais democrático na comunicação da literatura científica, como também a transparência nas avaliações dos trabalhos enviados às editoras científicas, divulgação da metodologia e dos instrumentos de coleta de dados, participação de pessoas externas à academia para o desenvolvimento de pesquisa científica, compartilhamento de dados abertos, entre outros aspectos (Cardoso; Jacobetty; Duarte, 2012).

Em 2009, Lucas e Willinsky, em uma pesquisa intitulada “*Open Access to e-Research*”, expõem limitações nas práticas exercidas pelas terminologias

empregadas para definir o fazer científico centrado em dados, a exemplo da e-*Science*. De acordo com os autores,

[...] a ambição expressa nessas frases tem sido até agora realizada apenas de forma incompleta. A e-*Science* tem feito muito para promover técnicas de criação de conhecimento, tais como a computação em grade e a colaboração remota em investigação, mas o seu envolvimento com questões de **disseminação tem sido mais limitado**. Foram feitos alguns avanços admiráveis na área da partilha de dados, mas foi dada menos atenção à **distribuição do conhecimento dentro da academia** e menos ainda à **circulação pública da investigação publicada** (Lucas; Willinsky, 2009, p. 270, grifo nosso).

Os autores, ao apresentar estas limitações da e-*Science* e de outros termos, possibilitam realizar uma análise das práticas da Ciência Aberta, que, diferentemente dos termos apresentados na seção anterior, promove um maior acesso às publicações e dados científicos, gerando benefícios tanto para os que fazem parte da comunidade científica quanto para aqueles que financiam este ecossistema com o pagamento de impostos (cidadãos). Nessa visão, a Ciência Aberta emerge na tentativa de suprir estas lacunas, com a incorporação de diversas iniciativas, principalmente dos movimentos de Acesso Aberto e Códigos Abertos.

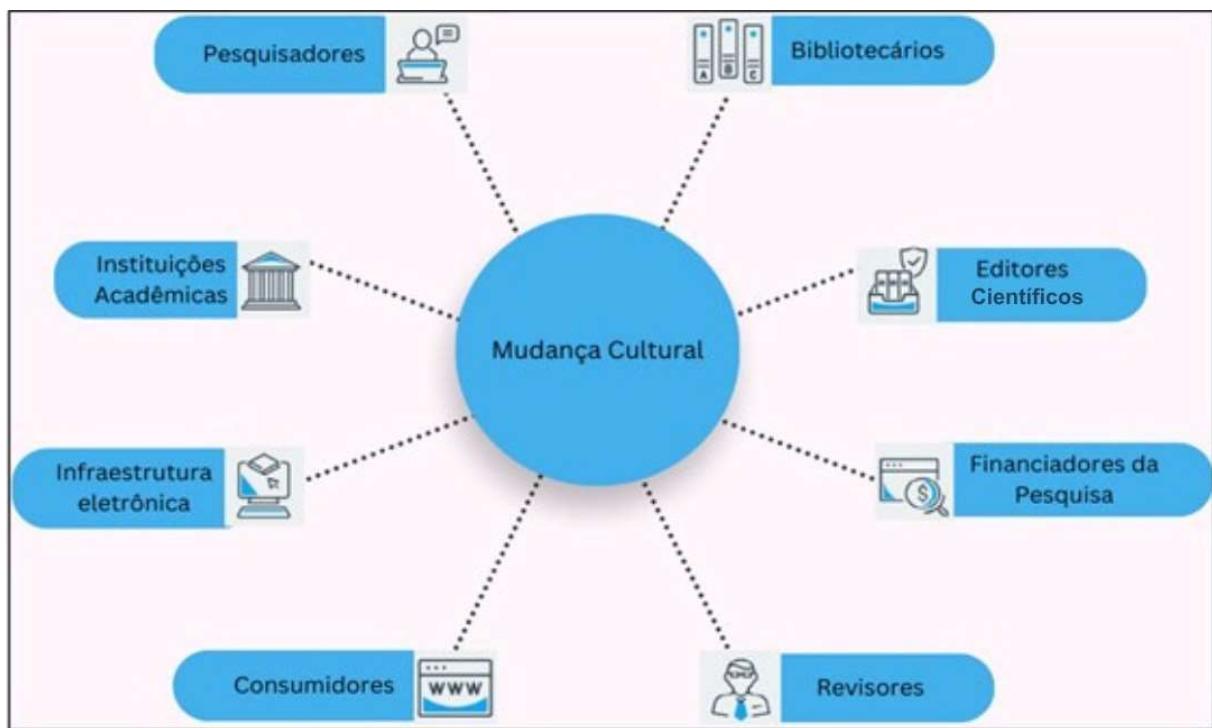
A ciência é feita para ser aberta e, de fato, beneficia-se quando não ocorrem restrições. Ela está descobrindo novas alternativas para que todos os envolvidos – comunidade científica e sociedade –, possam aproveitar, sem ter de pagar por isso novamente. Isso culmina em uma evidência já em prática: a Ciência Aberta deve ser o futuro (Silva; Silveira, 2019, p. 3).

Nesse sentido, a Ciência Aberta está difundida em diversas partes do mundo e seu alcance pode ser constatado nas iniciativas de organizações como *Center for Open Science* (COS), *GO FAIR* e *Open Knowledge Foundation* (OKF). Essas organizações criaram uma série de recomendações e ferramentas para tornar a ciência mais acessível, inclusiva, igualitária e sustentável. Destacam-se abaixo algumas iniciativas desses organismos.

A COS desenvolve práticas de pesquisa aberta no ambiente acadêmico com a finalidade de promover a democratização, integridade, autocorreção, transparência, compartilhamento, rigor e reproduzibilidade da pesquisa (COS, 2024). A comunidade

acadêmica idealizada pela COS fundamenta-se na participação de diversos atores e elementos para promover a mudança cultural no contexto acadêmico.

Figura 1 – Ecossistema científico da Center for Open Science



Fonte: elaboração própria com base na *Center for Open Science*².

Nesse ecossistema, cada elemento apresentado desempenha uma função fundamental nos incentivos à adoção de práticas da Ciência Aberta no ambiente das instituições acadêmicas, conforme o Quadro 1.

Quadro 1 - Mudança cultural da Center for Open Science

Infraestruturas eletrônicas	O conteúdo acadêmico é preservado, conectado e versionado para promover a descoberta, o acúmulo de evidências e o respeito a incerteza.
Instituições acadêmicas	Avaliam os pesquisadores com base no conteúdo de suas descobertas e no processo pelo qual foram descobertas, não em onde esses resultados são publicados.
Pesquisadores	Priorizam acertar em vez de publicar e receberão créditos por contribuições acadêmicas que vão além do artigo de pesquisa, como gerar dados úteis ou criar código que pode ser reutilizado por outros.

² Ver informações em: <https://www.cos.io/open-science>.

Bibliotecários	Aplicam conhecimentos de curadoria e gerenciamento de dados durante todo o ciclo de vida da pesquisa, não apenas retrospectivamente.
Editoras científicas	Competem com base na qualidade do serviço em vez de controlar o acesso ao conteúdo.
Financiadores da pesquisa	Têm total conhecimento dos retornos de seus investimentos em pesquisa para priorizar investimentos futuros.
Revisores	Fornecem <i>feedback</i> em todas as etapas do ciclo de vida da pesquisa e recebem crédito e melhoria de reputação pela revisão.
Consumidores	Têm acesso direto à revisão e à evidência primária para alegações acadêmicas.

Fonte: elaboração própria com base na *Center for Open Science*³.

Outra iniciativa de destaque da COS é a *Open Science Framework (OSF)*. Essa ferramenta tem como objetivo permitir e organizar os fluxos de trabalho das pesquisas para que se tornem mais colaborativos e transparentes, com foco na mudança cultural em direção à Ciência Aberta (COS, 2024). Entretanto, mudança cultural em um ambiente acadêmico é um processo complexo e demorado. Por esse motivo, a COS acredita que essa transformação pode ser iniciada por meio de estratégias de captação e treinamento de pessoas para que possam, posteriormente, servir de modelos na promoção de práticas da Ciência Aberta no contexto acadêmico. A mudança cultural aconteceria em um processo de dentro para fora.

A *OKF* foi fundada em 2004 e sua atuação é focada principalmente no desenvolvimento de redes de comunidades distribuídas em diversas partes do mundo (inclusive no Brasil), com foco no conhecimento aberto, criação de ferramentas digitais, cursos para capacitação, construção de políticas e metodologias que visam incentivar a abertura de informações e dados dos governos e instituições (OKF, 2024a). Para ajudar na promoção da Ciência Aberta pelo mundo, a *OKF* desenvolveu diversas ferramentas para cumprir esses objetivos, apresentadas no Quadro 2.

³ Ver informações em: <https://www.cos.io/open-science>.

Quadro 2 – Serviços e produtos ofertados pela OKF

Open Definition	Estabelece princípios que definem “abertura” em relação a dados e conteúdo. Ele torna preciso o significado de “aberto” nos termos “dados abertos” e “conteúdo aberto” e, assim, garante a qualidade e incentiva a compatibilidade entre diferentes conjuntos de material aberto.
Open Data Commons	Abriga um conjunto de ferramentas e licenças legais para ajudar a publicar, fornecer e usar dados abertos.
Ckan	Sistema de gerenciamento de dados de código aberto. O CKAN facilita a publicação, o compartilhamento e a descoberta de dados. Alimenta centenas de portais de dados no mundo todo, incluindo os de governos, ONGs, centros de pesquisa e empresas.
Frictionless Data	Aplicativo que permite trabalhar com dados complexos para organizá-los, uni-los, convertê-los e enviá-los para outros aplicativos, com a intenção de estabelecer análises ou armazenamentos automatizados. Aos pesquisadores, o aplicativo permite tornar os dados de pesquisa mais reprodutíveis; aos cientistas de dados, permite criar <i>pipelines</i> ⁴ de processamento de dados; e aos engenheiros de dados, permite padronizar plataformas de dados complexas.
Open Your Data	Resolve desafios técnicos, sociais e legais da abertura de dados. Trabalha para entender as necessidades, desenvolver estratégias de dados e estabelecer as melhores práticas.
Open Digital Public Infrastructure (DPI)	Orienta organizações na escolha de tecnologias abertas e padrões de dados.
School of Data	Oferece uma variedade de cursos gerais e personalizados sobre o uso de soluções abertas.

Fonte: elaboração própria com base na *Open Knowledge Fundation*⁵.

A GO FAIR Fundation é uma instituição que desenvolveu um conjunto de normas denominadas Princípios FAIR para permitir a descoberta e o gerenciamento dos dados de forma mais organizada. A sigla FAIR é um acrônimo do inglês formado pelas iniciais das palavras *Findable* (Encontrável), *Accessible* (Acessível), *Interoperable* (Interoperável) e *Reusable* (Reutilizável). Os Princípios FAIR têm como principal objetivo tornar os dados açãoáveis por máquina, “[...] ou seja, a capacidade

⁴ Ver informações em: <https://blog.dsacademy.com.br/mas-afinal-o-que-e-pipeline-de-dados/>

⁵ Ver informações em: <https://okfn.org/en/what-we-do/tools/>

dos sistemas computacionais de encontrar, acessar, interoperar e reutilizar dados com nenhuma ou mínima intervenção humana [...]” (GO FAIR, 2024), além de facilitar a reutilização dos dados pelos humanos.

Os Princípios FAIR possuem uma configuração bastante flexível, o que possibilita que cada componente da sua estrutura (Encontrável, Acessível, Interoperável e Reutilizável) possa ser utilizado de forma associada ou independente para conferir “[...] características que os recursos de dados, ferramentas, vocabulários e infraestruturas contemporâneas devem exibir para auxiliar a descoberta e a reutilização por terceiros” (Wilkinson, et al., 2016, p. 4).

Nesse sentido, seus mecanismos beneficiam pesquisadores, editores, agências de financiamento à pesquisa, criadores de softwares e ferramentas, que trabalham com mineração e gestão de dados. Outro fato interessante é que os Princípios FAIR não se restringem apenas ao uso para dados, mas também a todos os processos de uma pesquisa que levam até os dados, como, por exemplo, a obtenção de artigos científicos através de pesquisas. O Quadro 3 demonstra as recomendações criadas pela GO FAIR.

Quadro 3 – Princípios FAIR

FAIR	Descrição das recomendações
Findable	<p>F1. (meta)dados recebem um identificador globalmente único e persistente.</p> <p>F2. dados são descritos com metadados ricos (definidos por R1 abaixo).</p> <p>F3. os metadados incluem clara e explicitamente o identificador dos dados que descrevem.</p> <p>F4. (meta)dados são registrados ou indexados em um recurso pesquisável.</p>
Accessible	<p>A1. (meta)dados são recuperáveis por seu identificador usando um protocolo de comunicação padronizado:</p> <p>A1.1 o protocolo é aberto, gratuito e universalmente implementável;</p> <p>A1.2 o protocolo permite um procedimento de autenticação e autorização, quando necessário.</p> <p>A2. metadados são acessíveis, mesmo quando os dados não estão mais disponíveis.</p>
Interoperable	<p>I1. (meta)dados usam uma linguagem formal, acessível, compartilhada e amplamente aplicável para representação de conhecimento.</p> <p>I2. (meta)dados usam vocabulários que seguem os princípios.</p> <p>I3. (meta)dados incluem referências qualificadas a outros (meta)dados.</p>

Reusable	<p>R1. os metadados são descritos de forma rica com uma pluralidade de atributos precisos e relevantes.</p> <p>R1.1. os (meta)dados são liberados com uma licença de uso de dados clara e acessível.</p> <p>R1.2. os (meta)dados são associados a uma procedência detalhada.</p> <p>R1.3. os (meta)dados atendem aos padrões da comunidade relevantes para o domínio.</p>
-----------------	--

Fonte: Wilkinson *et al.* (2016).

Para difundir os Princípios *FAIR* pelo mundo, a *GO FAIR* trabalha com Redes de Implementação organizadas em torno de três pilares e com a participação de indivíduos, instituições e organizações (GO FAIR Foundation, 2024):

- *GO Change*: trabalha com mudança sociocultural através da construção de políticas e incentivos para a implementação dos Princípios FAIR e difusão da Ciência Aberta;
- *GO Train*: trabalha com capacitação e conscientização para construção de planos de gerenciamento de dados focados em Princípios FAIR;
- *GO Build*: trabalha com a construção e coordenação de infraestruturas eletrônicas para tornar os dados em *FAIR*.

No cenário brasileiro, a difusão e o desenvolvimento da Ciência Aberta estão fortemente atrelados às iniciativas de alguns organismos como Instituto Brasileiro de Informação, Ciência e Tecnologia (IBICT), Fundação Osvaldo Cruz (Fiocruz), Rede Brasileira para Educação e Pesquisa (RNP) e o Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). Para melhor compreensão da relevância desses institutos na comunicação científica brasileira no que tange à Ciência Aberta, é oportuno apresentar algumas informações sobre os projetos, eventos e ações que foram realizados nos últimos anos para o crescimento do Movimento no Brasil.

O IBICT foi fundado em 27 de fevereiro de 1954 com o nome de Instituto Brasileiro de Bibliografia e Documentação (IBBD); em 1976, passou a ser reconhecido com o seu nome atual e, no ano de 2000, tornou-se um órgão subordinado ao Ministério da Ciência e Tecnologia (IBICT, 2018).

Inspirado nas diversas declarações e manifestos emitidos para apoiar o acesso aberto pelo mundo, o IBICT publica, em 2005, o “Manifesto brasileiro de apoio ao acesso livre ao conhecimento científico”. O documento, escrito e proposto por Hélio

Kuramoto e Suely Costa, constitui-se em “[...] um conjunto de recomendações a todos os segmentos da comunidade científica para a implantação do acesso livre no país” (Kuramoto, 2009, p. 212), por meio de critérios objetivos como, por exemplo, o reconhecimento das publicações em acesso aberto para efeitos de avaliação e progressão dos docentes. O Manifesto é um convite aos principais responsáveis pela produção e disseminação da ciência no Brasil para que, juntos, contribuam e promovam acesso mais democrático à literatura científica no país.

O IBICT também estabeleceu projetos que consistiam em identificar *softwares* gratuitos e de código aberto desenvolvidos por instituições estrangeiras com a intenção de customizá-los e adaptá-los para a construção de repositórios no Brasil. Os *softwares* selecionados deveriam atender alguns critérios de interoperabilidade, como, por exemplo, suportar em seus sistemas o uso do *Open Archives Initiative Protocol for Metadata Harvesting* (OAI-PMH) e o conjunto de metadados *Dublin Core* (DC). Por meio dessas iniciativas, nasce o primeiro repositório do Brasil, a Biblioteca Digital Brasileira de Teses e Dissertações (BDTD) em 2002 (IBICT, 2023a; Kuramoto, 2009). Posteriormente, com o objetivo de promover o crescimento de repositórios no Brasil, foi elaborado o edital FINEP/PCAL/XBDB 002/2009 para contemplar algumas instituições de ensino e pesquisa com *kits* tecnológicos para construção de seus próprios repositórios institucionais e revistas eletrônicas (Kuramoto, 2009). Em 25 de novembro de 2009, o IBICT promoveu o I Encontro sobre Gestão de Repositório Institucional, com a presença das 27 instituições contempladas no edital FINEP/PCAL/XBDB 002/2009 e mais 6 instituições participantes do projeto-piloto para a implantação de repositórios, coordenado pelo IBICT⁶.

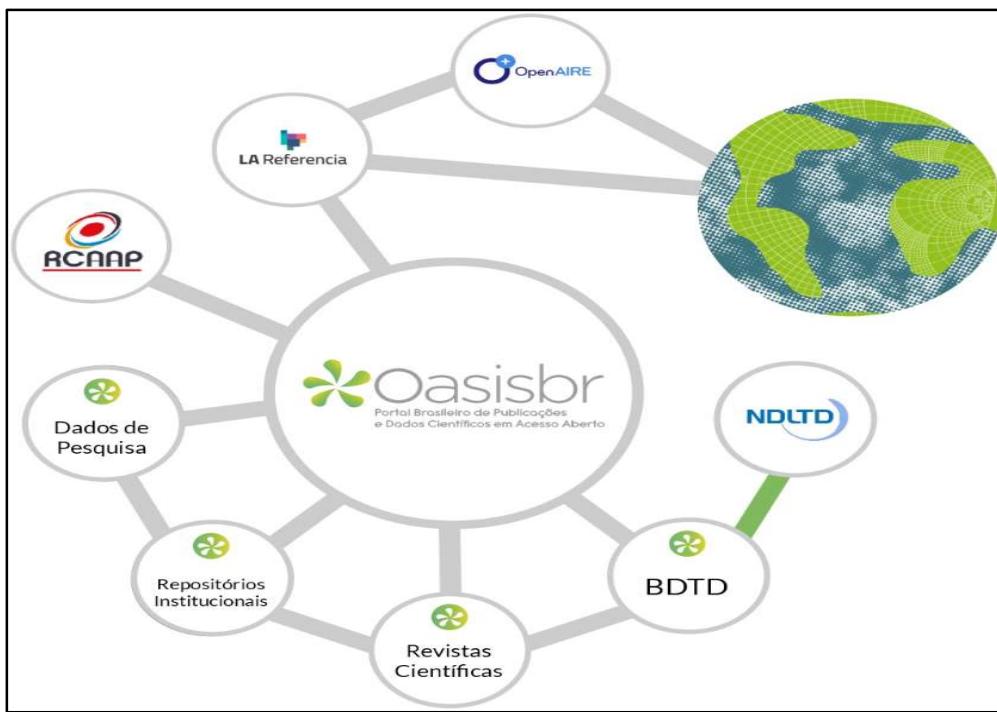
Em 2006, com um projeto de integração da produção científica nacional, o IBICT concebeu o *Open Access Scholarly Information System* (*Oasisbr*)⁷, que posteriormente passou a ser denominado de Portal Brasileiro de Publicações e Dados Científicos em Acesso Aberto. A plataforma coleta os metadados das dissertações, teses, artigos científicos, artigos de eventos, livros, capítulos de livros e conjuntos de dados de diversos repositórios institucionais e periódicos científicos do Brasil, possibilitando o acesso a essas produções em um único local, além de promover a

⁶ Para acessar mais informações, ver: <https://antigo.ibict.br/sala-de-imprensa/noticias/item/1302-sai-o-resultado-do-edital-finep-pcal-xbdb-002-2009>.

⁷ Ver informações em: <https://oasisbr.ibict.br/vufind/about/home>.

visibilidade tanto nacional quanto internacional dessas produções. A Figura 2 expõe o funcionamento do ecossistema de coleta de metadados do *Oasisbr*.

Figura 2 – Ecossistema da coleta de metadados do *Oasisbr*



Fonte: Oasisbr (2025⁸).

Em 2016, o IBICT lançou o “Manifesto de Acesso Aberto a Dados da Pesquisa Brasileira” e ratificou seu apoio à abertura e acessibilidade à produção científica brasileira, despondo como um dos principais órgãos no desenvolvimento da Ciência Aberta no Brasil. O principal foco desse Manifesto é promover o adequado gerenciamento dos dados científicos das pesquisas produzidas no país. Na visão do IBICT, os objetivos do Manifesto podem ser alcançados com responsabilidades atribuídas aos diferentes atores da disseminação e produção científica no Brasil, como: universidades e institutos de pesquisa brasileiros; sociedades científicas e academias de ciência; órgãos de fomento à pesquisa; editores de revistas ou periódicos científicos; cursos de pós-graduação e graduação nas áreas de informação, em especial Ciência da Informação; gestores e executores de programas e projetos de dados de pesquisa, especialistas em Ciência da Informação e Ciência da Computação e profissionais de informação em geral, pesquisadores.

⁸ Ver informações em: <https://oasisbr.ibict.br/vufind/>.

Destacamos ainda outras iniciativas criadas pelo IBICT que foram fundamentais para o desenvolvimento da Ciência Aberta no Brasil, conforme o Quadro 4: Conferência Lusófona de Ciência Aberta (ConfOA); Diretório de políticas editoriais de revistas brasileiras (Diadorim); Repositório Comum do Brasil (Deposita); *Emerging Research Information (EmeRI)*; Ecossistema Brasileiro de Informações para a Investigação Científica (BrCris); Portal Brasileiro para as revistas científicas (Manuelzão); Diretório de revistas científicas brasileiras (Miguilim); Repositório Comum de Dados de Pesquisa (Deposita Dados); Modelo de Preservação Digital Hipátia; Plataforma de Ciência Cidadã (Civis); Plano de Gestão de Dados do Brasil (PGD-BR); Rede Moara; e o *decentralized Archival Resource Key (dARK)* (Freitas-Campos *et al.*, 2023; IBICT, 2023).

Quadro 4 – Produtos, serviços e eventos criados pelo IBICT

ConfOA	“[...] tem como objetivo unir as comunidades de Portugal, Brasil e outros países de língua portuguesa que atuam nas áreas de pesquisa, desenvolvimento, gestão de serviços e formulação de políticas ligadas à Ciência Aberta, abrangendo aspectos como o Acesso Aberto à Informação Científica e Dados de Pesquisa e se posicionando como um espaço ideal para facilitar a troca, o debate e a divulgação de conhecimentos, práticas e estudos sobre esses temas, em suas várias dimensões e pontos de vista” (IBICT, 2024).
Diadorim	“é um serviço de informações relativas às autorizações concedidas para o armazenamento e o acesso dos artigos das revistas brasileiras em repositórios digitais de acesso aberto” (IBICT, 2025).
EmeRI	“O Repositório de <i>preprints</i> <i>Emerging Research Information (EmeRI)</i> foi implementado em 2020 a partir de uma cooperação entre a Associação Brasileira de Editores Científicos (ABEC) e o Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT) com o objetivo de prestar serviços às revistas e editores, de modo a agilizar a difusão de resultados de pesquisas científicas emergentes a partir da disponibilização de arquivos de <i>preprints</i> ” (IBICT, 2023).
BrCris	O Ecossistema de Informação da Pesquisa Científica Brasileira, BrCris, é uma plataforma agregadora que permite recuperar, certificar e visualizar dados e informações relativas aos diversos atores que atuam na pesquisa científica do contexto brasileiro (IBICT, 2025).
Manuelzão	“[...] visa contribuir com os editores científicos e equipes editoriais, uma vez que disponibiliza uma série de informações, produtos, serviços e orientações de como gerenciar as revistas e melhorar a disseminação e o impacto da produção científica nelas veiculadas. De forma mais específica, as informações disseminadas no Manuelzão têm como escopo quatro grandes áreas que consideram: revistas científicas e sua relação com a Ciência Aberta, a criação e editoração de revistas científicas, indexação e qualificação de revistas científicas e uma área

	que reúne conteúdos diversos de interesse para o editor científico (IBICT, 2025).
Miguilim	criada com o intuito de agregar, em um único local, informações sobre as revistas científicas editadas e publicadas no Brasil que se encontravam dispersas em diferentes plataformas. O Diretório reúne em sua base de dados o cadastro de informações essenciais das políticas editoriais de milhares de revistas científicas brasileiras [...]” (IBICT, 2025).
Deposita Dados	“[...] integra e dissemina conjuntos de dados de pesquisa de pesquisadores brasileiros vinculados a instituições científicas que ainda não possuem seus repositórios de dados de pesquisa e/ou de pesquisadores brasileiros que executaram seus conjuntos de dados por meio de colaboração científica em instituições estrangeiras de ensino e pesquisa. O acesso a esses conjuntos de dados é livre de quaisquer custos” (IBICT, 2025).
Deposita	“[...] é uma base de dados para reunir e divulgar a produção científica dos pesquisadores de instituições que ainda não possuem seus próprios repositórios institucionais. O Deposita tem por finalidade tornar mais acessível e visível a produção científica brasileira. A reunião e organização dessas publicações também objetiva promover a criação de repositórios institucionais em universidades e institutos de pesquisa que ainda não possuem este tipo de serviço” (IBICT, 2024).
Hipátia	é uma proposta de implementação de preservação digital arquivística a partir da integração dos ambientes de gestão, preservação e acesso que proporcione a garantia da autenticidade dos documentos ao longo do tempo. O modelo é baseado em normativas nacionais e internacionais que versam sobre as melhores práticas acerca da promoção da preservação digital, que também perpassam características de gestão e acesso (IBICT, 2025).
Cívus	“[...] é uma plataforma de Ciência Cidadã, desenvolvida pelo IBICT em código aberto a partir da <i>EU-Citizen Science</i> ⁹ . O “[...] objetivo é oferecer infraestrutura e conteúdos que contribuam para ampliar o entendimento sobre Ciência Cidadã, disseminar seu uso e dar suporte ao desenvolvimento de iniciativas e à aplicação de metodologias nesse campo, com foco na América Latina e no Caribe” (IBICT, 2025).
PGD-Br	“[...] é uma ferramenta de código aberto online gratuita que auxilia os(as) pesquisadores(as) brasileiros(as) a criarem seus Planos de Gestão de Dados (PGD)” (IBICT, 2025).
Rede Moara	“[...] é uma plataforma que reúne desenvolvedores de software livre, add-ons e interfaces de programação de aplicativos (API) relacionados à Ciência, Tecnologia e Inovação (CT&I)” (IBICT, 2025).
dARK	“um sistema descentralizado que permite a atribuição e gestão de identificadores com maior equidade e acessibilidade para todas as instituições de pesquisa, independentemente de seus recursos financeiros” (IBIC, 2024).

Fonte: elaboração própria com base nas páginas web do IBICT.

⁹ Ver informações em: <https://eu-citizen.science/>.

A Fiocruz é uma instituição de pesquisa fundada em 1900 que visa fortalecer o desenvolvimento social, por meio de pesquisas científicas voltadas principalmente para a descoberta de vacinas e a concepção de infraestruturas tecnológicas para a disseminação do conhecimento na área da saúde (Fiocruz, 2025). O instituto tem uma participação significativa para o desenvolvimento da Ciência Aberta no Brasil com a formulação de diversas iniciativas.

Em 2014, foi instituída a Política de Acesso Aberto ao Conhecimento da Fiocruz, para “[...] garantir à sociedade o acesso gratuito, público e aberto ao conteúdo integral da produção intelectual desenvolvida pela Fiocruz” (Fiocruz, p.1, 2014), assim como a preservação digital desse conteúdo. Para a materialização das diretrizes contidas na política, foi concebido o Repositório Institucional (ARCA) para o depósito dessa produção.

A partir de 2018, a Fiocruz começou a elaborar uma série de cursos sobre a Ciência Aberta na modalidade online, divididos em vários módulos, apresentados no Quadro 5. O objetivo visava promover a capacitação de pessoas nos diferentes movimentos que compõem o ecossistema da Ciência Aberta.

Quadro 5 – Cursos em Ciência Aberta da Fiocruz

Cursos	Conteúdo dos cursos
O que é Ciência Aberta?	1. Introdução à Ciência Aberta; 2. Acesso Aberto; 3. Dados de pesquisa abertos; 4. <i>Workflows</i> abertos; 5. Ciência Cidadã; 6. Inovação Aberta; 7. Educação Aberta; 8. Boas práticas e ferramentas de Ciência Aberta.
Panorama histórico da Ciência Aberta	1. Cenário Internacional; 2. Cenário Brasileiro; 3. Ciência aberta e saúde: abertura dos dados governamentais; 4. Os obstáculos que a Ciência Aberta pretende mitigar; 5. Uma ciência aberta, várias expectativas; 6. Visões críticas da Ciência Aberta. 7. Qual Ciência Aberta precisamos.
Propriedade intelectual aplicada à ciência aberta	Unidade 1. Propriedade Intelectual e Direitos Autorais 1. Propriedade Intelectual: Direitos Autorais; 2. História; 3. Direitos Autorais – Autoria; 4. Direitos Morais; 5. A obra protegida e não protegida; 6. Domínio público e limitações aos direitos autorais; 7. Transferência de Direitos Autorais. Unidade 2. Propriedade Industrial 1. Os desafios da inovação Aula; 2. Propriedade Intelectual: Propriedade Industrial; 3. As patentes e a criação do sistema internacional de proteção da Propriedade Industrial; 4. Para que servem e quais informações encontramos nas patentes?; 5. O modelo vigente da proteção das inovações; 6. Ciência Aberta – possibilidade de um novo modelo.

Direito de Acesso à Informação e Proteção de Dados Pessoais	Neste curso, os participantes terão noção da abrangência do direito de acesso à informação e poderão reconhecer as possibilidades de concessão ou negação de informações, em razão da publicidade ou sigilo dos dados com mais segurança jurídica, além de compreender a aplicação desses conceitos para a gestão de dados de pesquisa (Fiocruz, 2018).
Acesso Aberto	1. O que é Acesso Aberto?; 2. O panorama internacional; 3. A experiência nacional e regional; 4. Relatos da nossa experiência; 5. Implicações na Comunicação Científica.
Dados Abertos	1. Gestão, compartilhamento e abertura de dados para pesquisa: uma nova cultura no fazer científico; 2. Gestão “ativa” de dados: uma mudança cultural; 3. Dados e proteção jurídica; 4. Plano de Gestão de Dados; 5. FAIR: dos princípios à prática; 6. Políticas dos financiadores, requisitos das revistas científicas e repositórios de dados em saúde.
Educação Aberta	1. O que são Recursos Educacionais Abertos; 2. Como criar REA?; 3. Como encontrar, usar e compartilhar REA; 4. Monitoramento e avaliação da qualidade de REA.

Fonte: elaboração própria com base nos Cursos de Formação Modular sobre Ciência Aberta da Fiocruz¹⁰.

Posteriormente, em 2019, foi criada a plataforma de recursos educacionais abertos EDUCARE, “um ecossistema digital que oferece soluções para armazenar, disponibilizar e garantir o acesso de recursos educacionais abertos (REA) a toda sociedade” (Fiocruz, 2019). A plataforma foi concebida através de uma parceria entre a Fiocruz, a Organização Pan-Americana da Saúde (OPAS) e o Sistema Universidade Aberta do SUS (UNA-SUS) para gerenciar os recursos abertos educacionais produzidos pela Fiocruz.

Já em 2020, a Fiocruz elaborou a **Política de gestão, compartilhamento e abertura de dados para pesquisa**, que visa estabelecer a disponibilização dos dados provenientes das pesquisas desenvolvidas no âmbito da instituição através do Repositório ARCA Dados (Fiocruz, 2020). Este marco insere a Fiocruz em um cenário mundial que busca pavimentar o compartilhamento dos dados científicos produzidos pelas instituições de ensino e pesquisa.

A RNP foi fundada em 1989 com o objetivo de estabelecer, no âmbito acadêmico, uma infraestrutura nacional de rede de internet. A instituição, nos últimos

¹⁰ Ver informações em: <https://campusvirtual.fiocruz.br/portal/?q=formacao-modular/ciencia-aberta>.

anos, tem participado de forma ativa em diversos projetos para o desenvolvimento da Ciência Aberta no Brasil.

A Rede e-Ciência é uma infraestrutura eletrônica projetada pela RNP para atender centros de pesquisa, laboratórios multiusuários e instituições que trabalham com a geração de um grande volume de dados. A função da Rede é propiciar alto poder de processamento, segurança cibernética e rápida transferência de dados gerados nos projetos científicos de grande escala. A Rede é composta pelos seguintes membros: Laboratório Nacional de Computação Científica (LNCC), Senai Cimatec, Centro de Previsão de Tempo e Estudos Climáticos (CPTEC-INPE), Centro de Pesquisas da Petrobras (Cenpes), Centro Nacional de Pesquisa em Energia e Materiais (CNPEM), Centro Brasileiro de Pesquisas Físicas (CBPF), Empresa Brasileira de Pesquisa Agropecuária (Embrapa), Universidade Estadual Paulista (Unesp), Laboratório Multiusuário de Computação de Alto Desempenho da Universidade Federal de Goiás (LaMCAD/UFG) e Universidade Federal do Espírito Santo (Ufes) (RNP, 2025).

O *Alteryx Analytics* é um serviço ofertado pela RNP de forma gratuita aos pesquisadores, pós-graduandos e professores das instituições públicas que fazem parte da Rede CAFé¹¹. O serviço tem como objetivo o uso de ferramentas para proporcionar análise de dados, aplicação de técnicas de *machine learning* e uso de inteligência artificial, com foco na preparação de dados para a automatização de tarefas complexas, além criar soluções inovadoras no gerenciamento de dados (RNP, 2025).

O CNPq foi fundado em 1951 com a função de criar mecanismos e estratégias para o desenvolvimento da investigação científica e tecnológica no Brasil, principalmente com o financiamento de pesquisas científicas (CNPq, 2020). A instituição, no campo da Ciência Aberta, promoveu diversas ações para o desenvolvimento do Movimento no país.

No ano de 2020, o CNPq financiou a construção do Portal Ciência Aberta, que desenvolve projetos, infraestruturas eletrônicas de código aberto e o compartilhamento de pesquisas e dados para a promoção da Ciência Aberta no campo das Humanidades.

¹¹ Ver informações em: <https://aei.uefs.br/servicos-rnp/#:~:text=Rede%20CAFé,organiza%C3%A7%C3%B5es%20que%20participam%20da%20federal%C3%A7%C3%A3o>

O Portal Multimodal/Multilíngue pretende divulgar a Ciência Aberta na área das humanidades, não apenas compartilhando pesquisas e seus resultados, mas também disponibilizando e reunindo um conjunto de dados (verbais, visuais, verbo-visual, verbo-gestual) para que estudiosos tenham acesso a eles e os reutilizem em novas investigações, em diferentes áreas. Para tanto, esta plataforma fundamenta-se nos quatro princípios norteadores da Ciência Aberta: acesso aberto, dados abertos, fonte aberta e ciência cidadã (Portal Ciência Aberta, 2025).

O ano de 2022 é marcado pelo lançamento do Lattes Data, Repositório de dados científico do CNPq. A plataforma passa a ser o local em que serão reunidos, armazenados e disponibilizados os conjuntos de dados dos pesquisadores que recebem financiamento do CNPq. Existem planejamentos para que futuramente as instituições que fazem parte do Consórcio CoNCienciA possam também disponibilizar os seus conjuntos de dados no Lattes Data (Lattes Data, 2025).

Entre os projetos de parcerias interinstitucionais realizados pelo CNPq, destacamos a Rede de Dados de Pesquisa Brasileira (RDP Brasil), o Sistema de Informação sobre a Biodiversidade Brasileira (SiBBr) e o Consórcio Nacional para Ciência Aberta (CoNCienciA), descritos a seguir.

A RDP Brasil foi criada em 2017 através da parceria entre IBICT e RNP com a intenção de selecionar projetos de Grupo de Trabalho (GT) para colaborar com a identificação de práticas em Acesso Aberto em Dados de Pesquisa (AADP). O projeto dos pesquisadores da Universidade Federal do Rio Grande do Sul (UFRGS) foi selecionado para contribuir com a parceria (RDP Brasil, 2024).

O SiBBr foi fundado através da Portaria nº 6.223, de 29 de novembro de 2018, com parceria firmada entre o Ministério da Ciência, Tecnologia e Inovação (MCTI) e a RNP. A infraestrutura eletrônica no Brasil é responsável pela organização, compartilhamento e preservação de dados e informações sobre a biodiversidade e os ecossistemas brasileiros e faz parte do nó brasileiro na plataforma *Global Biodiversity Information Facility* (GBIF) (RNP, 2024).

Em 2021, iniciaram-se os planejamentos para criação do Consórcio Nacional para Ciência Aberta (CoNCienciA). O projeto visava apoiar a criação de repositórios de dados científicos no Brasil, assim como a obtenção do *Document Object Identifier* (DOI) a um baixo custo para atribuí-los aos conjuntos de dados depositados nos repositórios. O consórcio conta com a parceira do IBICT, Empresa Brasileira de

Pesquisa Agropecuária (Embrapa), Fiocruz e o Centro Brasileiro de Pesquisas Físicas (CBPF) (CNPq, 2022).

Importante destacar também outras instituições que já promoveram e promovem diversas iniciativas para o desenvolvimento da Ciência Aberta no Brasil: Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes), Associação Brasileira de Editores Científicos (ABEC), *Open Knowledge Foundation Brasil (OKF Brasil)* e a *Scientific Electronic Library Online (SciELO)*.

As ações, manifestos, eventos, cursos, serviços e infraestruturas eletrônicas criadas por estas instituições em prol da Ciência Aberta proporcionaram novos caminhos para o ecossistema científico brasileiro, conectaram o país com as inovações disponibilizadas pela comunidade internacional e trabalharam em prol da divulgação da produção do conhecimento científico do brasileiro em nosso país e no mundo.

Nessa perspectiva, as iniciativas dessas organizações evidenciam que a Ciência Aberta se tornou um dos principais aglutinadores de mudanças culturais do ecossistema científico. Todavia, é importante ressaltar que a Ciência Aberta não atua para modificar paradigmas científicos, sua atuação opera na perspectiva de tornar os processos de investigação e comunicação da ciência mais transparentes, acessíveis, colaborativos e inclusivos com a participação da sociedade (Cardoso; Jacobetty; Duarte, 2012; Silva; Silveira, 2019).

3.1 Ciência aberta: conceitos, características e taxonomias

A ciência é um campo de constantes quebras de paradigmas (Kuhn, 1975), sendo assim, conceituar determinada ação ou movimento científico é uma tarefa bastante complexa, ainda mais tratando-se de práticas que proporcionam grandes transformações, como é o caso da Ciência Aberta. Nessa perspectiva, a palavra “disruptiva” traduz significativamente a complexidade em torno da Ciência Aberta, pois trata-se de um movimento que rompe com práticas tradicionais da ciência à medida que engloba outros movimentos para a resolução de problemas comuns ao campo científico. Obviamente que essas características provocam divergências em seu campo conceitual por conta da participação de diversos interessados (Albagli, 2015; Vicent-Saez; Martinez-Fuentes, 2018).

A seguir, serão apresentadas algumas conceituações e características sobre a Ciência Aberta, mas sem a intenção de esgotar o tema ou apontar um conceito mais adequado, e sim com o objetivo de promover reflexões acerca da complexidade deste Movimento e proporcionar um aprofundamento da temática.

A definição para Ciência Aberta levantada por Silva e Silveira (2019) fundamenta-se em aspectos que evidenciam a transparência de todas as etapas dos processos que envolvem a pesquisa, divulgação e permissão de uso dos conteúdos científicos, com a participação de indivíduos que estão fora do contexto acadêmico. A visão dos autores entra em consonância com a definição do projeto *FOSTER* (2024), no qual a Ciência Aberta é reconhecida como um movimento em que “[...] outros podem colaborar e contribuir, onde dados de pesquisa, notas de laboratório e outros processos de pesquisa estão disponíveis gratuitamente, sob termos que permitem a reutilização, redistribuição e reprodução da pesquisa e seus dados”.

Para encontrar uma definição sobre Ciência Aberta, Vicent-Saez e Martinez-Fuentes (2018) desenvolveram uma pesquisa que analisou 75 estudos que abordam a temática. Os autores identificaram na literatura científica estudada que a palavra “conhecimento” aparece de forma recorrente para definir Ciência Aberta. Assim, a palavra “conhecimento” seria o “gênero”, a essência central do Movimento. Outras palavras que aparecem com menor recorrência, como transparência, acessível, compartilhamento e colaborativo, seriam os qualificadores desse Movimento. Desse modo, de acordo com os autores, “a Ciência Aberta é um conhecimento transparente e acessível que é compartilhado e desenvolvido por meio de redes colaborativas” (Vicent-Saez; Martinez-Fuentes, 2018, p. 434,).

Já conforme Cardoso, Jacobetty e Duarte (2012), a Ciência Aberta pode ser identificada como um movimento social criado por cientistas e que se manifesta como um contraponto a uma tendência crescente de restrição da publicação científica. Os autores ainda afirmam que a essência de movimento social da Ciência Aberta está ancorada em três pilares: 1) identidade: cientistas que partilham de um mesmo objetivo na busca de tornar a ciência mais acessível diante de obstáculos legais, econômicos e tecnológicos; 2) adversários: os que competem contra os objetivos da Ciência Aberta; 3) visão: como o movimento se define e como os seus objetivos e projetos se correlacionam para formação da sua identidade.

Em 2021, a Organização das Nações Unidas para a Educação, a Ciência e a Cultura (Unesco), formada por 193 países, elaborou uma série de recomendações para atualizar o campo de atuação da Ciência Aberta. As Recomendações da Unesco são considerada um marco no avanço do Movimento, pois, pela primeira vez, um organismo internacional com alcance global elabora um manifesto neste sentido e com estas proporções. O Manifesto evidencia a importância da adoção da Ciência Aberta pelos países membros da Unesco para o enfrentamento de desafios complexos que afligem o planeta. Na visão da organização, a Ciência Aberta é definida como:

[...] um construto inclusivo que combina vários movimentos e práticas que têm o objetivo de disponibilizar abertamente conhecimento científico multilíngue, torná-lo acessível e reutilizável para todos, aumentar as colaborações científicas e o compartilhamento de informações para o benefício da ciência e da sociedade, e abrir os processos de criação, avaliação e comunicação do conhecimento científico a atores da sociedade, além da comunidade científica tradicional. Abrange todas as disciplinas científicas e todos os aspectos das práticas acadêmicas, incluindo ciências básicas e aplicadas, ciências naturais, sociais e humanas [...] (Unesco, 2021, p. 7).

O Manifesto criado pela Unesco lança um novo olhar para a Ciência Aberta, pois elenca uma série de questões sobre a participação do movimento não somente para beneficiar o campo acadêmico e institutos de pesquisa, mas também para criar possibilidades de promover representatividade para grupos sociais historicamente excluídos, como mulheres, minorias étnicas e pesquisadores de países que estão fora do protagonismo científico, além de recomendar também o uso das práticas da Ciência Aberta para:

[...] enfrentar os complexos desafios ambientais, sociais e econômicos interligados às pessoas e o planeta, incluindo pobreza, questões de saúde, acesso à educação, aumento das desigualdades e disparidades de oportunidades, aumento das lacunas nas áreas de ciência, tecnologia e inovação, esgotamento dos recursos naturais, perda de biodiversidade, degradação da terra, mudança climática, desastres naturais e causados pelo homem, conflitos crescentes e crises humanitárias relacionadas [...] (Unesco, 2021).

Já Fecher e Friesike (2014), em pesquisa intitulada “*Open Science: one term, five schools of thought*”, optaram por não construírem um conceito para a Ciência Aberta, justificando que tal medida poderia provocar limitações nas discussões acerca

deste termo, ressaltando ainda que, apesar da confluência de ideias e objetivos, o termo também “evoca entendimentos bastante diferentes e abre uma infinidade de campos de batalha” (Fecher; Friesike, 2014, p. 17). Diante de tais justificativas, os autores construíram uma classificação em que a Ciência Aberta é organizada em cinco Escolas de Pensamento, e sua concepção foi extraída de diversas fontes de informação, como literaturas científicas, *blogs* científicos e artigos de jornais (Fecher; Friesike, 2014). A seguir, são descritas algumas características dessas escolas.

Escola Pública: é dividida em duas vertentes, em que a primeira é focada na promoção da participação de pessoas que estão fora do contexto acadêmico para contribuírem de forma ativa na construção de pesquisas científicas; a segunda é focada na reivindicação para a concepção de produções científicas com uma escrita que torne a ciência mais comprehensível para um público mais amplo. Nesse sentido, um dos principais objetivos dessa Escola é levar o conhecimento científico para um público que tradicionalmente não é inserido ou alcançado pelas produções da sociedade científica.

Escola Democrática: focada na acessibilidade da produção científica para qualquer pessoa, sem restrições legais, tecnológicas ou econômicas. A escola é dividida em duas vertentes: a *Open Access*, que é direcionada ao livre acesso das publicações científicas (livros e artigos científicos), e a segunda vertente, que é direcionada a disponibilização e reutilização dos dados científicos das produções científicas.

Escola Pragmática: focada no uso de tecnologias da *web 2.0* para tornar mais eficiente o processo da pesquisa e da sua disseminação, assim como o uso de ferramentas online para tornar a pesquisa mais colaborativa, com a participação de pesquisadores de diversas áreas do conhecimento para a resolução de problemas multidisciplinares.

Escola de Infraestrutura: focada no desenvolvimento de *software*, *hardware* e aplicativos de códigos abertos e não proprietários para que os pesquisadores possam produzir pesquisas colaborativas e possibilitar o compartilhamento do conhecimento científico de forma acessível. A Escola está dividida em duas vertentes que se correlacionam em muitos aspectos. A **computação distribuída**, que se manifesta em uma estrutura distribuída e interligada de diversos computadores de alta performance para desenvolver pesquisas colaborativas e complexas, principalmente

com o tratamento de uma imensa quantidade de dados. **Redes sociais e de Colaboração**, que promovem meios alternativos para que os pesquisadores possam discutir temas de interesse e promover pesquisas colaborativas com o uso da *web 2.0*.

Escola de Métricas: direciona diversas críticas ao modelo tradicional que mede o impacto das pesquisas científicas, pois muitos desses modelos, como o Fator de Impacto, por exemplo, visam mensurar o impacto dos periódicos ao invés do artigo científico, causando assim deturpações sobre a qualidade das pesquisas científicas. Os seguidores dessa Escola de Pensamento defendem a criação de modelos alternativos de medição dos impactos de uma pesquisa frente ao modelo tradicional. Um exemplo de medida alternativa seria a *Altmetrics*¹², que é formada por diferentes formas de medidas, como *blogs* científicos, marcação de favoritos, redes sociais científicas, *downloads* e compartilhamento de conjuntos de dados.

A intenção dos autores com a estruturação das cinco Escolas de Pensamento foca na identificação de padrões em torno das construções conceituais na Ciência Aberta e propõe novas possibilidades em relação ao seu futuro dentro do ecossistema científico. Um outro fator evidenciado na pesquisa discute que a diversidade e a ambiguidade no campo conceitual da Ciência Aberta são advindas da multiplicidade de partes interessadas, formadas por profissionais de TI, editores científicos, pesquisadores de diversos campos, governo (formuladores de políticas) e cidadãos que possuem interesse na temática (Fecher; Friesike, 2014).

Partilhando do mesmo ponto de vista de Fecher e Friesike (2014), a pesquisadora Albagli (2015, p. 11) caracteriza o Movimento fundamentando-se nos aspectos de sua complexidade conceitual e das atuações de diversos grupos que se interessam pelo movimento, assim a Ciência Aberta pode ser “[...] entendida como processo, algo em construção, que mobiliza interesses e pontos de vista distintos (e, em alguns aspectos, antagônicos); e que também permite múltiplas (e por vezes conflituosas) interpretações”. Dito isto, a Ciência Aberta é uma conjunção de vários movimentos com diversas ramificações: políticas para ciência aberta, dados científicos aberto, ciência cidadã, revisão aberta por pares, *softwares* e códigos-fontes abertos, *hardware* aberto, cadernos abertos de laboratório, pesquisa reproduzível aberta, recursos educacionais abertos, redes sociais científicas e dados abertos

¹² Ver informações em: <https://www.altmetric.com/>

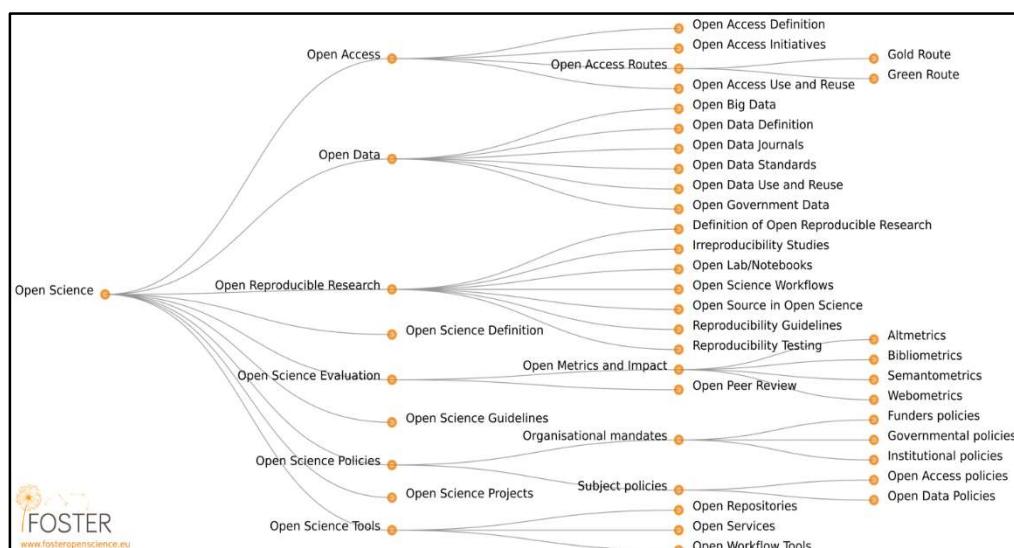
governamentais (Albagli; Clinio; Raychtock, 2014; Pontika *et al.* 2015; Silveira *et al.*, 2021; Unesco, 2021).

Diante das exposições conceituais apresentadas, a Ciência Aberta pode ser comparada a um movimento “antropofágico”, que, à luz de um dos modernistas de 1922, Oswald de Andrade, defende a importância de “devorar” o estrangeiro para criar a sua identidade brasileira. Desse modo, à medida que a Ciência Aberta incorpora outros movimentos em seu escopo, os seus propósitos, alcance e ações são ressignificados, obtendo novas características para o seu complexo ecossistema.

Para exemplificar a proporção das ramificações que a Ciência Aberta alcançou, pesquisadores desenvolveram modelos de taxonomias para representar as influências e atuações do Movimento. Segundo Silveira *et al.* (2023), “uma taxonomia é uma classificação sistemática e hierárquica de conceitos, objetos ou eventos de acordo com as suas características ou propriedades comuns”.

Uma das taxonomias mais conhecidas no meio científico foi criada por diversos pesquisadores do grupo *FOSTER* (*Facilitate Open Science Training for European Research*). A taxonomia descreve como a Ciência Aberta é formada por outros movimentos e como esses movimentos e suas subdivisões estão interligados (Figura 3).

Figura 3 – Taxonomia da Ciência Aberta *FOSTER*



Fonte: Pontika *et al.* (2015).

A taxonomia proposta pela *FOSTER* é composta por nove ramificações primárias que representam os principais movimentos que integram a Ciência Aberta, denominadas, respectivamente, de Acesso Aberto, Dados Abertos, Pesquisa Reprodutível Aberta, Definição de Ciência Aberta, Avaliação da Ciência Aberta, Diretrizes de Ciência Aberta, Políticas de Ciência Aberta, Projetos de Ciência Aberta e Ferramentas de Ciência Aberta, posteriormente ramificadas em vinte e quatro rótulos secundários e mais onze terciários.

Consequentemente, com o surgimento de novos desafios tecnológicos, sociais e políticos que atravessam a ciência e influenciam suas práticas de produção e comunicação, outros pesquisadores propuseram expandir o ecossistema da Ciência Aberta. O objetivo era remodelar a antiga taxonomia para um contexto científico mais atual e provocar novas reflexões em relação às atuações da Ciência Aberta. Neste sentido, é importante destacar duas pesquisas desenvolvidas com participação de pesquisadores brasileiros publicadas em 2021 e 2023.

A primeira pesquisa foi composta por dez pesquisadores de diferentes universidades federais brasileiras com a proposta de atualizar a taxonomia criada pelo grupo *FOSTER*. Nesta perspectiva, houve mudanças significativas na antiga taxonomia, como: a inclusão de cinco novas ramificações primárias interligadas a Ciência Aberta denominadas de Educação Aberta, Licenciamento Aberto, Ciência Cidadã, Preservação Digital e Inovação Aberta; entre as ramificações secundárias, terciárias e quaternárias foram criados quarenta e sete novos rótulos; o termo **Definição de Ciência Aberta** foi excluído pelo entendimento de que já estava contemplado em outros termos; o termo **Diretrizes de Ciência Aberta** passou a integrar o termo **Políticas de Ciência Aberta**, que, por sua vez, manteve sua mesma nomenclatura, e o termo **Projetos de Ciência Aberta** foi substituído pelo termo **Ciência Cidadã** pelo entendimento de ser mais representativo ao contexto (Silveira et al., 2021). A representação da taxonomia pode ser visualizada no Anexo 1 deste estudo.

A segunda pesquisa contou com a participação de treze pesquisadores, sendo seis do Brasil, três da Colômbia, dois da Costa Rica, um da Espanha e um do Peru. Propuseram realizar um estudo comparativo das taxonomias *FOSTER* (2015), Baumgartner (2019) e Silveira et al. (2021), tomando como base o relatório “Recomendações da UNESCO sobre Ciência Aberta” (Unesco, 2021). O estudo

também pretendia apresentar uma taxonomia que contemplasse a pesquisa científica no contexto da América Latina e superasse uma visão sobre a Ciência Aberta exclusivamente formada por perspectivas do Norte Global (Europa, Estados Unidos e Canadá). Ao término da pesquisa, chegou-se às seguintes mudanças: composição de dez ramificações primárias interligadas a Ciência Aberta; adição de outra ramificação primária, **Diálogo aberto com outros Sistemas de Conhecimento**; exclusão de **Licenciamento aberto** como ramificação primária e sua incorporação como um rótulo dentro da ramificação primária **Acesso Aberto**; exclusão de **Preservação digital** como ramificação primária e sua incorporação como um rótulo dentro das ramificações primárias **Infraestrutura e Ferramentas e Dados abertos**, e, ao todo, foram criados noventa e seis rótulos (Silveira *et al.*, 2023). A representação da taxonomia pode ser visualizada no Anexo 2.

A construção dessas taxonomias é importante para ilustrar o crescimento e a atualização da Ciência Aberta com a formação de diversos movimentos e suas ramificações, contudo, o mais importante no contexto dessas construções é a capacidade de proporcionar direcionamentos e informações aos *stakeholders* da Ciência Aberta (pesquisadores, editores, agências de fomento à pesquisa, profissionais de TI, bibliotecários e instituições de ensino e pesquisa) na formulação de políticas, métodos, estratégias, capacitações e construções de novas ferramentas de tecnologia (Fecher; Friesike, 2014; Silveira, *et al.*, 2023).

As taxonomias, porém, ainda carecem de uma representatividade ilustrativa no que tange aos relacionamentos e colaborações entre os principais movimentos dentro do ecossistema da Ciência Aberta. Um dos grandes exemplos é a influência da filosofia de abertura e compartilhamento do **Acesso Aberto** em movimentos como **Dados Abertos**, **Pesquisa Reprodutível Aberta** e **Avaliação da Ciência Aberta**. Nessa mesma abordagem, como pensar a concepção de outros movimentos sem as infraestruturas eletrônicas construídas no âmbito do movimento de **Ferramentas de Pesquisa Aberta**? Nesta contextualização, o ecossistema da Ciência Aberta assemelha-se ao funcionamento de um organismo humano, em que cada órgão desempenha uma função específica com relações dependentes e, ao mesmo tempo, interdependentes.

A seguir, no Quadro 6, estão dispostos conceitos sobre alguns movimentos da Ciência Aberta com o objetivo de proporcionar informações sobre seus significados e

suas atuações. Ressaltando que os movimentos de Acesso Aberto e Dados abertos não estão presentes nesta tabela, pois serão abordados com maior profundidade nas subseções 3.2, 3.3 e 4.1.

Quadro 6 – Movimentos do Ecossistema da Ciência Aberta

Movimentos	Conceitos e definições
Pesquisa Reprodutível Aberta/Cadernos Abertos de Laboratório	<p>“um modo de fazer ciência em que - da melhor maneira possível - você torna a sua pesquisa aberta ao público em tempo real” (Bradley, 2010¹³).</p> <p>“É uma ferramenta organizacional e auxiliar de memória, que serve como o registro primário de pesquisa e atividade científica para todos os cientistas. Ele também serve como um registro legal de propriedade das ideias e resultados obtidos por um cientista” (Schnell, 2015, p.1).</p>
Avaliação da Ciência Aberta	<p>Segundo a <i>Surrey of University</i>¹⁴ do Reino Unido esse processo pode ser compreendido através de diferentes modalidades como:</p> <p>Identidades abertas: as identidades do autor e do revisor são divulgadas (entre si e para o leitor)</p> <p>Relatórios abertos: Os relatórios de revisão são publicados juntamente com o artigo</p> <p>Participação aberta: A comunidade em geral pode contribuir para o processo de revisão</p> <p>Interação aberta: discussões diretas entre autores, editores e revisores são incentivadas.</p> <p>Revisão por pares aberta de pré-impressão: a revisão de pré-impressões (por exemplo, em servidores de pré-impressão) é possível antes da revisão por pares formal</p> <p>Comentários pós-publicação: os leitores podem comentar; e os autores/outros leitores respondem após a publicação</p> <p>Plataformas abertas (revisão dissociada): a revisão é facilitada por uma entidade organizacional diferente da editora.</p>
Políticas de Ciência Aberta	[...] são documentos normativos de países, instituições de pesquisa e agências de fomento que

¹³ Ver informações em: Entrevista de Jean-Claude Bradley concedida a Richard Poynder. Disponível em: <https://www.infotoday.com/it/sep10/Poynder.shtml#top>. Acesso em: 20 jan. 2025.

¹⁴ <https://www.surrey.ac.uk/library/open-research/open-peer-review>

	<p>promovem as práticas da ciência aberta (Costa, 2019¹⁵).</p>
Ferramentas de Ciência Aberta/Códigos Abertos	<p>“incluem softwares cujo código-fonte é disponibilizado publicamente, de forma oportuna e de fácil utilização, em formato modificável e legível por humanos e máquinas, sob uma licença aberta que concede a outros o direito de usar, acessar, modificar, expandir, estudar, criar trabalhos derivados e compartilhar o software e seu código-fonte, desenho ou projeto. O código-fonte deve ser incluído na versão do software e disponibilizado em repositórios de acesso aberto, e a licença escolhida deve permitir alterações, trabalhos derivados e compartilhamento em termos e condições abertas iguais ou compatíveis” (Unesco, 2021, p. 10).</p>
Educação Aberta	<p>[...] materiais de ensino, aprendizagem e investigação em quaisquer suportes, digitais ou outros, que se situem no domínio público ou que tenham sido divulgados sob licença aberta que permite acesso, uso, adaptação e redistribuição gratuitos por terceiros, mediante nenhuma restrição ou poucas restrições. O licenciamento aberto é construído no âmbito da estrutura existente dos direitos de propriedade intelectual, tais como se encontram definidos por convenções internacionais pertinentes, e respeita a autoria da obra” (Declaração REA de Paris, 2012).</p>
Ciência Cidadã	<p>[...] refere-se ao envolvimento do público em geral em atividades de pesquisa científica quando os cidadãos contribuemativamente para a ciência, seja com seu esforço intelectual ou conhecimento circundante ou com suas ferramentas e recursos. Os participantes fornecem dados experimentais e facilidades para pesquisadores, levantam novas questões e co-criam uma nova cultura científica. Ao agregar valor, os voluntários adquirem novos aprendizados e habilidades, e uma compreensão mais profunda do trabalho científico de forma atraente. Como resultado desse cenário aberto, em rede e transdisciplinar, as interações ciênciasociedade-política são aprimoradas, levando a uma pesquisa mais democrática baseada na tomada de</p>

¹⁵ Ver informações em: chrome-extension://efaidnbmnnibpcajpcgclefindmkaj/https://www.gov.br/capes/pt-br/centrais-de-conteudo/20191113_2EncontroCAPESdeCienciaAberta_MichelleCosta.pdf

	decisões informadas por evidências" (<i>Socientize Consortium</i> , 2013, p. 6, tradução nossa).
Licenciamento Aberto	"[...] especifica o que pode e o que não pode ser feito com uma obra (seja som, texto, imagem ou multimídia). Ela concede permissões e estabelece restrições. Em termos gerais, uma licença aberta é aquela que concede permissão para acessar, reutilizar e redistribuir uma obra com poucas ou nenhuma restrição" (OKF, 2024).

Fonte: elaboração própria.

A Ciência Aberta desponta no meio científico como um dos principais instrumentos para tornar os processos de pesquisa e comunicação mais abertos, acessíveis, colaborativos, reproduzíveis e transparentes. Isso gera impactos positivos não somente entre a comunidade científica, mas também em toda a sociedade. Entretanto, para que a Ciência Aberta não se transforme em um discurso utópico e distante nos países emergentes, principalmente no Brasil, que nos últimos anos sofreu com diversos contingenciamentos financeiros na área da ciência, há uma urgência na formulação de políticas, capacitação de pessoas e investimentos financeiros de ordem governamental.

3.2 Movimento de Acesso Aberto *versus* oligopólio editorial

O modelo de negócio exercido pelos grandes grupos editoriais possibilitou a mercantilização do conhecimento científico e impactou o comportamento da comunidade científica, levando à criação do Movimento de Acesso Aberto. O movimento é considerado por alguns teóricos como uma das principais ferramentas para a formulação dos alicerces da Ciência Aberta. Segundo Bertin *et al.* (2019, p. 3), "[...] no Brasil, pode-se afirmar que os primórdios da Ciência Aberta evoluíram do próprio movimento pelo Acesso Aberto à informação científica [...]", e as suas influências permeiam grande parte da taxonomia da Ciência Aberta, como já foi demonstrado na seção anterior. Nessa perspectiva, a seção é destinada a apresentar o percurso revolucionário e questionador do Movimento de Acesso Aberto nos mecanismos de produção e comunicação do ecossistema científico.

No decorrer do século XX, a comunidade científica chancelou os periódicos científicos como um dos principais canais formais para estabelecer a comunicação científica. Este formato de disseminação da informação surge 300 anos antes, no

século XVII, para atender aos anseios da comunidade científica por um instrumento regular e rápido para a distribuição das novas descobertas (Meadows, 1999; Ziman, 1979).

Com o advento das Tecnologias de Informação e Comunicação (TICs), a comunidade científica conseguiu potencializar o alcance das divulgações de suas pesquisas através dos periódicos científicos eletrônicos, resultando em grandes transformações no processo de produção e obtenção da informação científica. Os periódicos científicos eletrônicos tornaram a comunicação científica ainda mais dinâmica. Contudo, o uso das TICs para alavancar o processo da comunicação científica foi apropriado por grupos editoriais, seguindo o mesmo *modus operandi* realizado no formato impresso dos periódicos científicos.

A comunidade científica esperava que o uso das TICs provocasse mudanças significativas no ecossistema científico, com a criação de condições favoráveis para que a informação fosse distribuída de forma mais acessível e colaborativa. No entanto, essa lógica foi subvertida, pois a comunidade científica está inserida em uma sociedade capitalista e os efeitos dessa sociedade reverberam nas estruturas da sua comunidade também (Castells, 1999; Mueller, 2006). Com isso, grupos editoriais passaram a definir as formas de acesso ao conhecimento científico.

Um grande exemplo foi o advento do periódico científico eletrônico. Compreendia-se que este mecanismo, pelo fato de eliminar algumas etapas no processo da editoração – como por exemplo a impressão mecânica – proporcionaria um acesso menos custoso aos artigos científicos. Porém, os grandes grupos editoriais se beneficiaram dessa nova modalidade de disseminação científica para diminuir os seus custos e potencializar os seus lucros. O cenário torna-se ainda mais problemático quando se percebe que grande parte da publicação científica se encontra sob o poder de um pequeno grupo de editoras científicas, mas com grandes poderes decisórios no ecossistema da comunicação científica (Larivière; Haustein; Mongeon, 2015).

Por ser estruturado, o sistema de produção científica sofre de males profundos, pois não garante que o conhecimento seja genuíno ou que a legitimação do conhecimento produzido seja igual à sua qualidade; restringe o acesso ao conhecimento e favorece seu controle por poucos; responde, em muitos casos, aos interesses das grandes corporações; divide a sociedade entre ‘quem sabe’ (especialistas) e

quem ‘não sabe’ (outros); e, por fim, não inclui todos os atores envolvidos nos processos científicos (Silva; Silveira, 2019, p. 8).

A pesquisa desenvolvida por Larivière, Haustein e Mongeon (2015), sobre a existência de um oligopólio na indústria editorial científica, permite entender a dimensão da apropriação da produção intelectual no campo científico para obtenção de lucro. Para exemplificar o tamanho desse oligopólio, os autores afirmam que entre 2006 e 2013, a porcentagem dos lucros de algumas editoras chegou a superar a margem de empresas como a *Pfizer* e *Hyundai Motors*. Em 2011, editoras como a *Reed Elsevier*, *Springer*, *Taylor & Francis*, e *John Wiley & Sons* obtiveram, juntas, uma receita anual de 9,4 bilhões de dólares americanos. O estudo ainda revela que, em 2013, a *Reed Elsevier*, *Springer Science+Business Media*, *Wiley Blackwell*, *Taylor & Francis* e *Sage* foram responsáveis por 50% de todos os artigos indexados na *Web of Science*, sendo que, na área das ciências sociais, este número chegou a 70% (Larivière; Haustein; Mongeon, 2015).

O sucesso da consolidação desse oligopólio, entre outros fatores, pode ser atribuído à retenção dos direitos autorais para a permissão de publicação de artigos nestes periódicos. Assim, ao passo que o autor cede seus direitos autorais para promover a disseminação dos resultados de suas pesquisas e atender às avaliações em torno de sua carreira acadêmica/científica, o pesquisador e a instituição produtora deste conhecimento não possuem mais escolha sobre o acesso dessa produção científica (Barreto Segundo *et al.*, 2019). A comunidade científica está diante de um cenário em que a máxima estabelecida é “publicar ou perecer”, e a cessão dos direitos autorais acabou se tornando uma prática comum ao pesquisador.

Nesse contexto, a retenção dos direitos autorais é utilizada como um instrumento para que os grandes oligopólios passem a decidir a forma de distribuição e acesso à produção científica que deveria ser comunal (Dantas, 2014; Merton, 1973; Larivière; Haustein; Mongeon, 2015).

Como consequência desse oligopólio, os custos para poder acessar as produções científicas estavam ficando cada vez mais onerosos e provocavam uma “marginalização científica” de muitos pesquisadores e instituições que não tinham recursos financeiros para custear este mecanismo. Assim, as práticas estabelecidas pelo oligopólio editorial acabaram desencadeando a crise dos periódicos científicos, que teve o seu auge na década de 1980.

O gatilho da crise foi a impossibilidade de as bibliotecas universitárias e de pesquisa americanas continuarem a manter suas coleções de periódicos e a corresponder a uma crescente demanda de seus usuários, impossibilidade decorrente da falta de financiamento para a conta apresentada pelas editoras, cada ano mais alta, mais alta mesmo que a inflação e outros índices que medem a economia (Mueller, 2006, p. 31).

O modelo de negócio praticado nos anos de 1980, e ainda vigente pelos conglomerados editoriais para permitir o funcionamento do ecossistema da produção científica, consiste no pagamento de taxas para aceite da submissão de artigos científicos e uma outra taxa para disponibilizar o acesso a este. O cenário demonstra um modelo de negócio bastante peculiar, já que “os editores não pagam o fornecedor do bem primário – os autores de artigos acadêmicos [e científicos]. Nem o controle de qualidade – a revisão por pares” (Larivière; Haustein; Mongeon, 2015, p. 11-12). Na prática, vivencia-se uma contradição, pois grande parte da produção científica é subsidiada com investimentos governamentais e, mesmo assim, as publicações resultantes das pesquisas tornam-se posse de grupos editoriais.

Como exemplo desta contradição, podemos destacar o caso do Brasil, onde o Estado é o grande responsável pela manutenção do ensino superior. O ciclo envolve o uso de recursos financeiros públicos para a construção de Institutos de Ensino Superior (IES), formação de pesquisadores, disponibilização de agências de fomento à pesquisa e o financiamento dos pesquisadores para a disseminação dos resultados de suas pesquisas em diferentes canais de comunicação, a exemplo dos periódicos científicos. Ao término de todo esse processo, o Estado disponibilizará recursos financeiros para ter acesso aos resultados das pesquisas disseminadas em diferentes canais de comunicação, que em todo o seu ciclo produtivo foram financiadas por ele mesmo (Mueller, 2006).

Neste mecanismo, o Estado brasileiro ajuda a financiar um oligopólio que exerce uma cultura de poder fundamentada no uso da produção intelectual para a obtenção de lucro. Na estrutura desse mercado editorial, “[...] o capital só existe e subsiste com base na apropriação e privatização do produto de trabalho abstrato” (Dantas, 2014, p. 52).

Um outro dado importante a se destacar é que, no Brasil, quase a totalidade dos artigos científicos é oriunda de pesquisas realizadas por pesquisadores de instituições públicas. Segundo o levantamento feito pela *Clarivate Analytics* a pedido

da Capes, no período de 2011-2016, essa totalidade representa 95% da produção científica nacional, revelando-se um processo contraditório, pois grande parte destas produções possuem restrições para seu acesso. Contrastando com este modelo, Merton (1973) discorre que o conhecimento científico é um bem imaterial que deveria ser entendido através de uma ótica comunal, que pertence a toda a sociedade e a sua existência deveria ser concebida para o desenvolvimento da humanidade.

Esse cenário de oligopólio editorial científico impulsionou a organização, por parte da comunidade científica, para a construção de novos caminhos e práticas que possibilitaram um acesso mais democrático à informação científica. De acordo com Mueller (2006), essa organização da comunidade científica resultou no Movimento de Acesso Aberto no final do século XX.

Segundo Rosa e Gomes (2010, p. 22), o Movimento de Acesso Aberto ou Livre pode ser entendido como “a disponibilização na internet da literatura acadêmica e científica permitindo que seja lida, descarregada, distribuída, impressa, pesquisada contribuindo para o avanço e disseminação da ciência”. Esta literatura acadêmica e científica compreende as teses, dissertações, artigos científicos, artigos de eventos, relatórios de pesquisa, livros e capítulos de livros entre outros documentos.

As recomendações contidas no Movimento de Acesso Aberto têm a intenção de fortalecer uma rede colaborativa de troca de informações científicas, ao mesmo tempo que se contrapõem às práticas estabelecidas pelo oligopólio editorial. A premissa encontra sustentabilidade na visão de Meadows (1999, p. 49), a qual afirma que “os resultados de novas pesquisas dependem, em última instância, da interação entre cientistas e devem, por sua vez, ser colocados à disposição da comunidade científica”. Entretanto, para a disseminação ampla e democrática no campo científico, faz-se necessário o acesso aberto, visto que esta prática se configura quando não existem barreiras legais, técnicas ou financeiras para o acesso à produção científica.

Apesar do movimento passar a existir de forma mais estruturada no início do século XXI, já havia algumas ações para promover acesso mais democrático ao conhecimento. Entre algumas ações, podemos citar o *Project Gutenberg*¹⁶, de 1971, fundado por Michael Hart, considerado o criador do e-book – o projeto tinha como objetivo disponibilizar livros eletrônicos gratuitos; a criação da revista *Psycoloqui*, de

¹⁶ Ver informações em: <https://www.gutenberg.org/about/>

acesso livre, em 1989, e o lançamento da Proposta Subversiva, de 1995, ambos idealizados por Stevan Harnd (Harnard, 1995) e o Repositório *arXiv*¹⁷, lançado por Paul Ginsparg em 1991 na área de física, um dos pioneiros na disponibilização livre de conteúdo acadêmico. Atualmente, o *arXiv* possui diversos trabalhos em diferentes áreas do conhecimento, com mais de 2 milhões de arquivos.

No Brasil, destacamos as ações da *Scientific Electronic Library Online* (*SciELO*), que na década de 1990 passou a incentivar que as revistas científicas disponibilizassem os artigos em sua base de dados sem restrições de acesso. A iniciativa proporcionou novos caminhos para visibilidade no contexto da publicação científica para os pesquisadores no Brasil e em outros países (Shintaku; Ferreira Júnior, 2020).

3.2.1 Declarações, manifestos e infraestruturas eletrônicas do Movimento de Acesso Aberto

Os esforços isolados para facilitar o acesso à produção científica, mencionados na seção anterior, passam a ganhar outro *status*, à medida que a comunidade científica se organiza para a criação de documentos normativos fundamentados na promoção do acesso livre à literatura científica e na construção de infraestruturas eletrônicas de código aberto. Nesse contexto, diversos manifestos, declarações e cartas são emitidos através das reuniões que aconteceram em diversas partes do mundo. Assim, no final do século XX e início do século XXI, são emitidas quatro declarações consideradas basilares para o aprimoramento do Movimento de Acesso Aberto: *The Santa Fe Convention for the Open Archives Initiative* (1999), *Bethesda Statement on Open Access Publishing* (2003), *Berlin Declaration An Open Access to Scientific Knowledge* (2003) e *Budapest Open Access Initiative* (2002) – respectivamente traduzidas no Brasil como Declaração de Santa Fé, Declaração de Bethesda, Declaração de Berlim e Declaração de Budapeste (Marcondes; Sayão, 2009; Rios; Lucas; Amorim, 2019).

A Declaração de Santa Fé foi um manifesto criado a partir das reuniões realizadas em 21 e 22 de outubro de 1999, na cidade de Santa Fé, México. As reuniões contaram com a participação de cientistas da informação e bibliotecários

¹⁷ Ver informações em: <https://arxiv.org/>

digitais (Van de Sompel; Lagoze, 2000). O objetivo era criar recomendações para a construção de *softwares* de código aberto para autoarquivamento da produção dos autores acadêmicos nessas infraestruturas eletrônicas e construção de protocolos e conjuntos de metadados abertos para transpor as barreiras existentes na comunicação de *softwares* com infraestruturas distintas, entre outras iniciativas¹⁸.

A Declaração de Bethesda define características para a publicação em acesso aberto, com ênfase na área de Ciência Biomédica, ao mesmo tempo que ressalta a importância da participação de pesquisadores, instituições de pesquisa, agências financeiradoras, editores, sociedades científicas e bibliotecas para o alcance de suas propostas.

A Declaração de Berlim, em relação à de Bethesda, acrescenta nas suas recomendações os arquivos, museus e associações profissionais, também como responsáveis pela alteração do processo de difusão do conhecimento. Enfatiza a necessidade do uso de licenças para orientar os utilizadores no uso e atribuição de autoria das publicações.

A Declaração de Budapeste propõe dois métodos para eliminar as barreiras anteriormente existentes no acesso à produção científica, amplamente difundidos em vários países, nomeados de “vias”. A “Via Dourada”, com recomendações para que as universidades e institutos de pesquisa passem a utilizar estruturas eletrônicas para a construção e o gerenciamento do fluxo de publicação de periódicos científicos sem restrições no acesso de seus conteúdos, e a “Via Verde”, com recomendações para que as universidades e institutos de pesquisa utilizem estruturas eletrônicas para o autoarquivamento das suas produções acadêmicas, com o objetivo de colocá-las em acesso aberto.

Nesse contexto, podemos citar ainda outras declarações que, em algum aspecto, também contribuíram para ampliar o alcance do Movimento de Acesso Aberto em diferentes áreas do conhecimento e diversas partes do mundo, apresentadas no Quadro 7.

Quadro 7 – Declarações do Movimento de Acesso Aberto

Declarações	Ano	Objetivos

¹⁸ Ver informações em: <https://www.openarchives.org/>.

Princípios das Bermudas	1996	Os princípios afirmam que toda a informação da sequência genômica humana, gerada por centros financiados para sequenciação humana em grande escala, deve estar disponível gratuitamente e ser de domínio público.
Declaração de São José	1998	Construção de forma cooperativa da Biblioteca Virtual em Saúde (BVS), como estratégia de gestão da informação e do conhecimento em saúde para a Região Latino-Americana e do Caribe, com expansão para Espanha, Portugal, Moçambique e outros países de idioma português.
Declaração de Montreal (<i>Declaration on Free Access to Law</i>)	2002	Reconhece a informação jurídica pública de todos os países e instituições internacionais como parte do património comum da humanidade. Sendo assim, devem ser aplicados esforços para que seja acessível a todos, sem fins lucrativos e gratuitamente pela internet.
Declaração de Valparaiso	2004	Possui dez recomendações com melhorias para a comunicação científica em meio eletrônico com foco no acesso aberto.
Declaração da IFLA (<i>the International Federation of Library Associations and Institutions</i>)	2004	Está empenhada em garantir o acesso mais amplo possível à informação para todos os povos e apresenta duas definições para o reconhecimento de uma publicação em acesso aberto.
Declaração de Viena	2005	São apresentadas medidas e ações necessárias a serem tomadas para permitir o melhor acesso possível à informação por meio de recomendações intituladas “10 teses sobre liberdade de informação”.
Declaração de Salvador	2005	Reconhecer o acesso aberto como ferramenta de melhoria para a investigação científica, comunicação científica, educação e intercâmbio da ciência entre países. Reivindicação da implementação do acesso aberto nas políticas científicas desenvolvidas pelos países.
Declaração de Brisbane	2008	Reconhecer o acesso aberto como uma atividade estratégica facilitadora, na qual a investigação se baseia a nível internacional, nacional, universitário, de grupo e individual.

Fonte: elaboração própria com base na Declaração de São José, Declaração de Montreal, Declaração de Valparaiso, Princípios de Bermudas, Declaração da IFLA, Declaração de Viena, Declaração de Brisbane e Declaração de Salvador.

A materialização das recomendações presentes nos manifestos e declarações que foram apresentadas só foram possíveis com o surgimento de *softwares* de código aberto denominados repositórios. O uso dos repositórios pelas universidades e institutos de pesquisa é inspirado principalmente nas recomendações da Declaração de Budapeste (Via Dourada e Via Verde) e da Convenção de Santa Fé. A função destas estruturas eletrônicas é reunir, organizar, preservar e permitir o acesso à produção científica de uma determinada instituição ou área do conhecimento (Leite, 2009). A seguir, será apresentada algumas informações de dois *softwares* (*Open Journal System* e *DSPACE*) mais utilizados para este propósito.

O *Open Journal System* (OJS) é um *software* bastante utilizado para construção de repositórios para atender às recomendações da “Via Dourada”. O *software* foi desenvolvido pelo *Public Knowledge Project* (PKP), com sede na Universidade de Stanford. O projeto conta com quatro parceiros: Conselho de Bibliotecas Universitárias de Ontário, as Bibliotecas da Universidade de Alberta, as Bibliotecas da Universidade de Columbia Britânica e as Bibliotecas da Universidade de Pittsburgh. Segundo os dados do Instituto Brasileiro de Informação Ciência e Tecnologia (IBICT, 2023b) e do *Public Knowledge Project* (2021), mais de trinta mil periódicos utilizam o *software* e mais de oito milhões de artigos já foram publicados na sua plataforma em 150 países. A infraestrutura do OJS permite a viabilização da construção de portais de periódicos institucionais e temáticos, ou seja, a concepção de diversos periódicos em um único ambiente digital. Essa infraestrutura proporciona ações centralizadas de atualização do *software* e do gerenciamento de todo o fluxo editorial de um periódico científico: submissão, avaliação, edição de texto, editoração e publicação (Bandeira, 2017; IBICT, 2023b; *Public Knowledge Project*, 2021; Universidade Federal do Paraná, 2023).

Para atender as recomendações da “Via Verde”, existe uma grande variedade de *softwares* de códigos abertos disponíveis para construção de repositórios institucionais ou temáticos, porém, o mais utilizado atualmente é o *DSPACE*. Este *software* foi desenvolvido pela *Massachusetts Institute of Technology* (MIT) e a *Hewlett-Packard* (HP). Atualmente, a *DuraSpace* é a responsável pelo seu desenvolvimento e promoção pelo mundo (IBICT, 2023a). Para demonstrar a grande prática do seu uso, entre os 5.842 repositórios cadastrados no diretório do OpenDOAR (2024), um dos maiores diretórios de repositório do mundo, 40% dos repositórios

utilizam o *DSPACE*. O *E-prints* é o segundo mais utilizado, com 10%, e o *software Weko* aparece em terceiro, com 9% de uso.

As iniciativas da comunidade científica visando o crescimento do Acesso Aberto não ficaram restritas apenas às ações mencionadas anteriormente. Os anos que se seguiram foram de intensa realização de eventos, desenvolvimento de políticas e construção de produções (artigos científicos, livros, teses e dissertações), que fomentaram discussões acerca das restrições às produções científicas.

Como exemplo dessa contínua luta, recentemente, em 2022, foi lançada a *BOAI20* em comemoração ao 20º aniversário da Declaração de Budapeste de 2002. A *BOAI20* é uma declaração que reafirma os compromissos do Comitê Diretor da *BOAI* com os novos desafios que se apresentam à comunidade científica nos investimentos, produção, avaliação e comunicação dos seus trabalhos. A sua elaboração é baseada na primeira Declaração de Budapeste de 2002 e está dividida em quatro pontos principais com recomendações para os próximos dez anos (Budapeste [...], 2022):

1. Hospedar pesquisas de Acesso Aberto sobre infraestrutura aberta;
2. Reformar a avaliação e as recompensas da investigação para melhorar os incentivos;
3. Favorecer canais de publicação e distribuição inclusivos que nunca excluam autores por razões econômicas;
4. Quando utilizar dinheiro para publicar pesquisas sobre acesso aberto, lembre-se dos objetivos para os quais o acesso aberto é o meio (Budapeste [...], 2022).

Apesar dos diversos avanços ao longo dos últimos anos na luta pela democratização do acesso à informação no meio científico, a comunidade científica reconhece que ainda existem diversas barreiras presentes no ecossistema científico para o desenvolvimento mais amplo do acesso aberto.

4 A IMPORTÂNCIA DOS DADOS NA SOCIEDADE CONTEMPORÂNEA

A etimologia da palavra “dado” é proveniente do latim (*datum*) e suas origens remontam aos meados do século VII, advinda do campo filosófico (Semidão, 2014). De acordo com o Dicionário Michaelis (2024¹⁹), o seu significado é definido como aquilo “que foi concedido ou ofertado”. Segundo esta mesma vertente, Buckland (1991, p. 355) define dados como “coisas que podem ser dadas”, pensamento que segue uma abordagem em sentido etimológico da palavra.

Já na interpretação de Silva, Schwitzer e Rezende (2024, p. 206), o elemento é conceituado em sua forma singular (dado); assim, “dado é a menor unidade cognitiva e que sua compreensão depende de sua estruturação em informação para poder ser apropriada por indivíduos ou máquinas”, suscitando a ideia de algo em estado bruto, um elemento inicial, que ainda precisa ser processado, interpretado e contextualizado para revelar sua importância e aplicação de seu uso. Esta perspectiva se manifesta em diversas esferas da sociedade, que são exemplificadas a seguir.

A teoria intitulada “Pirâmide da Informação” ou “Hierarquia da Informação”, representada pelo acrônimo em inglês *Data, Information, Knowlegde e Wisdom* (DIKW), é uma construção conceitual que evidencia o dado como elemento inicial para a transformação do cognitivo humano (Sharma, 2016). A mudança cognitiva começa com a obtenção do dado, possibilitando, assim, desencadear acesso a outros níveis, tais como informação, conhecimento e sabedoria. A autoria do conceito DIKW é um tanto imprecisa na literatura e o seu surgimento sugere uma contribuição coletiva de diversos autores ao longo dos anos, passando pelo discurso do administrador Russell Ackof em 1989, pelo artigo futurista “*Information as Resource*” do cientista da informação Harland Cleveland em 1982, e pelo poema “*The Rock*” de T.S. Eliot em 1934 (Sharma, 2016).

No campo da Ciência da Informação, a recorrência da terminologia “dados” em seu arcabouço teórico pode ser entendida como componente de uma “tríade”, representada por dados, informação e conhecimento (Semidão, 2014; Souza; Almeida, 2021). O conceito de “tríade” aqui apresentado aponta os dados como elemento inicial para obtenção de novas ideias, segue o raciocínio de que os “[...] dados constituiriam o patamar menos dotado de significação. Informação surge a

¹⁹ Ver informações em: <https://michaelis.uol.com.br/busca?r=0&f=0&t=0&palavra=etimologia>.

partir da reunião e agregação de sentido aos dados. E o conhecimento, como patamar mais alto, culmina o trajeto com a assimilação do significado" (Semidão, 2014, p. 11). Assim, a "tríade" manifesta-se diacronicamente em diversos estudos dos principais teóricos da Ciência da Informação, como, por exemplo, Borko (1968), Buckland (1991), Saracevic (1999) e Le Coadiac (2004).

Ciência da informação é um campo interdisciplinar que comprehende várias disciplinas. Assim, o foco de estudo em ciência da informação é a informação e as revoluções causadas pelo uso e reuso da tecnologia de dados digitais. Dados digitais são insumos para novos sistemas tecnológicos; dados digitais são uma nova forma de informação composta pelo comportamento humano de dados mediados e gerados pelo uso intensivo de dados de tecnologias necessárias para as atuais atividades humanas de trabalho, saúde, educação, segurança, esportes, etc (Semeler; Pinto, 2020, p. 3).

Corroborando com as concepções apresentadas anteriormente, Silva (2019) afirma que os dados são o ponto fulcral para o desenvolvimento de um sistema cíclico de *input* e *output* que proporciona diversas possibilidades para produção de novas ideias. Esse ciclo se manifesta na seguinte ordem: "Dados – Informação – Conhecimento – Informação – Dados" (Silva, 2019, p. 23).

No tocante ao mundo corporativo, os dados tornaram-se uma das principais peças de destaque para as empresas que almejam alcançar os primeiros lugares em seus negócios, e assim tem sido para muitas empresas como a *Alphabet* (proprietária do Google), *Meta* (proprietária do Facebook, Instagram e WhatsApp), *Apple*, *Amazon*, entre outras. O grande volume de dados (*Big Data*) gerados pelos diversos produtos e serviços que estas empresas oferecem são rigorosamente coletados, tratados e estruturados para gerar vantagem competitiva nas tomadas de decisões em seus negócios (Taurion, 2013).

Analizando ainda o mundo corporativo e sua relação com os dados, apresenta-se a frase proferida pelo matemático britânico Clive Humby em 2006, afirmando que "os dados são o novo petróleo" (University of Sheffield, 2024). A frase é uma referência às transformações desencadeadas no mundo pela importância alcançada pelos dados, assim como foi com o petróleo algumas décadas atrás. O paralelo entre dados e petróleo, criado por Clive Humby, também pode ser entendido como uma analogia aos semelhantes processos de mineração, refino e transformação do produto principal

em subprodutos pelos quais o petróleo e os dados passam (*University Of Sheffield*, c2024).

A frase proferida por Clive Humby em 2006, acerca da emergente importância dos dados, torna-se ainda mais significativa perante o escândalo com os dados vazados de 87 milhões de usuários do Facebook pela *Cambridge Analytica*²⁰. O escândalo foi levado ao conhecimento público em 11 de dezembro de 2015, após as investigações jornalísticas de Harry Davies, do jornal inglês *The Guardian* (Fornasier; Beck, 2020). A prática da *Cambridge Analytica* consistia no uso de um aplicativo denominado *thisisyoudigitalife* para realizar coletas ilegais de diversos dados pessoais e sensíveis dos usuários do Facebook e comercializá-los com terceiros. Os dados foram utilizados para a produção de vídeos e notícias para influenciar a opinião de diversos eleitores na votação do *Referendum do BREXIT*²¹ em 2015 e nas eleições dos Estados Unidos da América (EUA) em 2016 (Fornasier; Beck, 2020; Solon, 2018). O escândalo da *Cambridge Analytica* é uma demonstração da relevância que os dados alcançaram nas últimas décadas, podendo interferir até em escolhas para a presidência de um país.

A importância dos dados também está atrelada ao desenvolvimento econômico dos países. No Brasil, por exemplo, na área da agricultura, um setor que foi responsável por 23,8% do Produto Interno Bruto (PIB) do país em 2023, o processo da coleta e tratamento de dados envolve drones, satélites, robôs autônomos, grandes bases de dados agrícolas e profissionais com conhecimento aprofundado no gerenciamento desses dados (Tilman *et al.*, 2002; Centro de Estudos Avançados em Economia Aplicada; Confederação da Agricultura e Pecuária do Brasil, 2024). Posteriormente, os dados coletados são aplicados para ajudar na correção da acidez do solo, irrigações automatizadas em áreas mais necessitadas, controle de pragas, novas cultivares e mecanização da colheita para a seleção das melhores safras, assim, promovendo um aumento significativo na produção de diversas culturas (Tilman *et al.*, 2002).

No campo da saúde, como exemplo da importância e uso dos dados, podemos destacar a pandemia do novo vírus Sars-CoV-2. O gerenciamento dos dados referentes à crise sanitária foi crucial para o combate da covid-19. Nesse contexto,

²⁰ Ver informações em: <https://www.investopedia.com/terms/c/cambridge-analytica.asp>

²¹ Ver informações em: <https://www.gov.br/empresas-e-negocios/pt-br/invest-export-brasil/exportar/conheca-os-mercados/brazil-brexit-watch/o-que-e-o-brexit>

foram criados diversos mecanismos em todo o mundo para coletar, tratar e cruzar os dados, com a intenção de monitorar e criar estratégias de enfrentamento à pandemia. No Brasil, o Instituto de Comunicação e Informação em Saúde (Icict/Fiocruz) criou o MonitoraCovid-19, uma ferramenta que agrupa e cruza dados sobre o novo coronavírus no Brasil e no mundo. A ferramenta disponibiliza diversos dados estruturados para o monitoramento da covid-19 no Brasil, além de permitir a comparação do avanço da doença com outros países.

Em nível mundial, foi implementada a *Virus Outbreak Data Network (VODAN)* em 2020, sob o comando da GO FAIR, com a parceria da CODATA, Research Data Alliance (RDA), World Data System (WDS), com a participação de dezenove países, incluindo o Brasil. O projeto consistia em criar uma infraestrutura de dados federada, com a inserção dos dados dos prontuários médicos de pacientes que se contaminaram com o Sars-CoV-2 ou outros vírus de grande contágio. Posteriormente, aplicaram-se os Princípios FAIR para tornar os dados inseridos no sistema em Localizáveis, Acessíveis, Interoperáveis e Reutilizáveis por humanos e máquinas (Veiga *et al.*, 2021). A infraestrutura VODAN possuía uma grande vantagem perante outros projetos, pois, além de permitir a identificação de padrões de surtos pandêmicos do Sars-CoV-2 para aplicar medidas de enfrentamento, também permitia acessar todos os dados do percurso do tratamento do paciente até a sua alta clínica, uma informação valiosa para futuros estudos clínicos (Veiga *et al.*, 2021).

Uma outra importância do uso dos dados, ainda se tratando do surto pandêmico do Sars-CoV-2, pode ser evidenciada com a criação de infraestruturas eletrônicas para disponibilizar o acesso aos dados de pesquisas científicas sobre o vírus. Assim, os esforços para o sequenciamento genético do Sars-CoV-2 para a produção de vacinas foram acelerados, além de possibilitar novos caminhos para estudos referentes ao Sars-CoV-2 e outros vírus. O fato evidenciou que as recomendações da Ciência Aberta podem proporcionar diversos avanços para o campo científico e, como consequência, beneficiar toda a sociedade. Nesse contexto, a cultura de restrição à produção científica teria que ser repensada.

4.1 Dados científicos

O relatório da *International Data Corporation (IDC)*, de 2017, prevê que, em 2025, a quantidade de dados gerados pelo uso da internet alcançará a marca de “[...]

163 zettabytes (ou seja, um trilhão de gigabytes). Isso representará dez vezes os 16,1ZB de dados gerados em 2016" (Reinsel; Gantz; Rydning, 2017, p. 3), um aumento de mais de 1000%, sendo que, desses dados, 20% serão críticos e 10% hipercríticos para a viabilização do desenvolvimento de empresas e da vida profissional de muitas pessoas. O processo de manipulação dessa imensa quantidade de dados exigirá um planejamento criterioso, profissionais altamente capacitados e aparatos tecnológicos cada vez mais sofisticados.

Foi durante a segunda metade do século XX que a percepção da importância dos dados como ponto de partida para diversas aplicações experimentou uma crescente em seu uso. Consequentemente, em decorrência deste fato, foram desenvolvidas tecnologias para produção e gestão destes dados que alcançaram números nunca vistos, principalmente com o advento da internet, como descrito anteriormente.

No campo científico também não foi diferente: nas últimas décadas, a comunidade científica vivenciou novos desafios para o desenvolvimento de infraestruturas ancoradas em *hardwares* e *softwares* específicos e potentes para proceder com a captura, análise, curadoria e compartilhamento dos dados científicos. Mas, afinal, o que são dados científicos?

O conceito apresentado pela *Organisation for Economic Co-Operation and Development (OCDE)* (2007, p. 13) define os dados científicos "como registros factuais (pontuações numéricas, registros textuais, imagens e sons) usados como fontes primárias para pesquisa científica que são comumente aceitos na comunidade científica como necessários para validar resultados da pesquisa". Nesse sentido, os dados são uma das principais fontes de informação para que os pares do campo científico possam verificar as etapas percorridas durante a pesquisa, com objetivo de identificar possíveis manipulações em seu resultado.

Na análise de Sayão e Sales (2016, p. 94), "quase tudo que é gerado e coletado no ambiente de pesquisa pode ser considerado dado de pesquisa", ressaltando que os dados são gerados das mais diferentes formas e os valores agregados a este insumo apresentarão conceituações diversificadas, sendo o seu acesso fundamental para o desenvolvimento de outras pesquisas.

Para Silva (2019, p. 21), "os dados científicos são as informações registradas ou produzidas através de qualquer forma ou meio durante o decurso de uma

pesquisa". Isso leva a compreensão de que, a depender da área do conhecimento e por quais processos esses dados foram gerados, teremos diversas tipologias de dados como: números, questionários, imagens, avaliações, textos, registros de casos, vídeos, áudio, protocolos de estudo, *software*, banco de dados, planilhas, algoritmos, notas de laboratório, equações, notas de campo, animações, sequências genéticas, modelos, metodologias, protocolos, diários, simulações e coordenadas geográficas (Sayão; Sales, 2015; Semeler; Pinto, 2019; Silva, 2019).

Por meio dos conceitos apresentados anteriormente, podemos conhecer algumas das características dos dados científicos: fontes primárias de pesquisa, validação da pesquisa, diferentes processos de construção e tipologia diversificada.

A geração de dados apresenta uma outra característica, evidenciada a depender da disciplina que foi gerada, analisada do ponto de vista dos experimentos desenvolvidos nas “*big sciences*” e nas “*small sciences*”. As duas terminologias personificam o conceito da “Cauda Longa” aplicado para o campo científico (Borgman, 2016). No conceito da “Cauda Longa da Ciência”, as “*big sciences*” são a cabeça da cauda, sendo formada por áreas do conhecimento como astronomia, oceanografia, biologia e física. Na outra extremidade, no final da “Cauda Longa”, temos as “*small sciences*”.

Nas áreas conhecidas como “*big sciences*”, o contexto científico influenciado pelas novas práticas advindas do “quarto paradigma” possibilitou uma cultura mais colaborativa entre diferentes áreas do conhecimento, culminando com surgimento de novos campos de estudos interdisciplinares, como biologia computacional e astroinformática (Borgman, 2012). Esse novo cenário, como consequência, desencadeou um dilúvio de dados no campo científico, entre outros fatores.

As “*big sciences*” geralmente são caracterizadas por grande destinação de investimento financeiro do setor público e pelo desenvolvimento de pesquisas a longo prazo que produzem uma imensa quantidade de dados que podem alcançar números em torno de petabytes a exabytes (Borgman, 2012, 2016; Sayão; Sales, 2019). Para exemplificar a quantidade de dados que são gerados nas “*big sciences*”, podemos citar dois exemplos: o projeto do Grande Colisor de Hádrons (GCH)²² da *European Organization for Nuclear Research* (CERN) e a técnica denominada Microscopia Eletrônica na área da Neurobiologia.

²² Ver informações em: <https://home.cern/science/accelerators/large-hadron-collider>

O GCH está localizado na fronteira entre a França e a Suíça e tem o formato de um anel com 27 km de circunferência que fica a 100 metros de profundidade em relação à superfície. O projeto é responsável pela geração de mais de 30 petabytes de dados por ano e mais de 100 petabytes estão armazenados permanentemente no *CERN Data Center*. Para desenvolver toda essa infraestrutura, o projeto conta com a participação de cientistas de diversos países. Na microscopia eletrônica²³, os dados do mapeamento de 1mm cúbico do tecido neural correspondem a 1 petabyte de dados, e o cérebro humano é composto por quase 1 milhão de mm cúbicos de tecido neural, impondo desafios de grandes proporções aos pesquisadores para gerir esses dados (Lichtman *et al.*, 2009). Nessa mesma perspectiva de projetos com grandes produções de dados, poderíamos citar ainda o *Pan-STARRS*²⁴; *Human Pangenome Project*²⁵; *James Webb Space Telescope*²⁶ e *NEPTUNE*²⁷.

Nas áreas denominadas “*small sciences*”, as características predominantes são de projetos científicos de curta duração, executados por pesquisadores de forma individual ou por pequenos laboratórios e com poucos dados gerados em relação às “*big sciences*”. Contudo, “[...] os dados da cauda longa **quando somados** representam a maior parcela de dados produzidos pela ciência [...]” (Sayão; Sales, 2019, p. 34, grifo nosso), além de uma produção bastante diversificada das tipologias de dados. É nesta esfera da ciência que geralmente existe um recorrente desconhecimento de informações referentes à gestão de dados e carência de verba, o que acaba culminando com uma grande perda de insumos valiosos para a ciência.

O termo “Cauda Longa da Ciência” é um conceito que propicia diversos estudos dos campos científicos em relação à sua caracterização na perspectiva do fazer científico centrado em dados. Entretanto, “a metáfora da ‘cauda longa da ciência’ é problemática por muitas razões, entre as quais se destaca o facto de reduzir a rica complexidade da ciência a duas dimensões” (Borgman, 2016, p. 129). Essa complexidade pode ser percebida em muitas disciplinas, como astronomia ou biologia, que são consideradas “*big sciences*”, e mesmo assim, existem o desenvolvimento de pequenos projetos e pesquisas individuais. Um outro fator é que os termos “*big*

²³ Ver informações em: <https://afinkopolimeros.com.br/o-que-e-microscopia-eletronica-mev/>

²⁴ Ver informações em: <https://www2.ifa.hawaii.edu/research/Pan-STARRS.shtml>

²⁵ Ver informações em: <https://humanpangenomeproject.org/>

²⁶ Ver informações em: <https://science.nasa.gov/mission/webb/>

²⁷ Ver informações em: <http://ooicruises.ocean.washington.edu/enlighten/story/The+NEPTUNE+Concept>

sciences" e "*small sciences*" podem produzir interpretações que determinadas áreas do campo científico possuem mais importância do que outras para o desenvolvimento da ciência.

Corroborando com a visão de Borgman (2016), para fins desta pesquisa, o termo Grandes Projetos Científicos (GPC) será adotado em substituição ao termo "*big sciences*" e o termo Pequenos Projetos Científicos (PPC) será adotado em substituição ao termo "*small sciences*", com a intenção de evitar interpretações inadequadas. É importante ressaltar também que o objeto da presente pesquisa é voltado para a gestão dos dados científicos no campo dos Pequenos Projetos Científicos, realidade bastante recorrente nos Institutos de Ensino Superior (IES).

5 BIBLIOTECAS UNIVERSITÁRIAS NO SUPORTE À GESTÃO DE DADOS CIENTÍFICOS

As recomendações para compartilhar os dados científicos são um processo recente no âmbito das IES no Brasil, e seus critérios envolvem informações e procedimentos que, geralmente, são desconhecidos ou parcialmente conhecidos pelos pesquisadores. Portanto, é natural que novas práticas provoquem resistência em sua adesão, principalmente na esfera dos Pequenos Projetos Científicos (PPC), onde os recursos para o desenvolvimento de pesquisa, via de regra, são mais escassos.

O não compartilhamento e a resistência ao compartilhamento dos dados científicos no âmbito das IES podem estar atrelados a diversos fatores, como: questões éticas relacionadas ao plágio; necessidade de diferentes abordagens no tratamento dos dados; receio que outros pesquisadores encontrem inconsistências em suas pesquisas durante a reproduzibilidade; falta de política ou serviços institucionais para orientar o pesquisador (Shintaku; Ferreira Júnior, 2021), variações de formatos de arquivos e tipologias de dados, resultando em um entendimento mais complexo desse processo, pois “dados qualitativos por exemplo, possuem singularidades que requerem atenções diferentes dos quantitativos” (Shintaku; Ferreira Júnior, 2021, p. 232), entre outros fatores.

Um outro problema recorrente nesse contexto é a forma como esses conjuntos de dados são armazenados, muitas vezes em suportes inadequados ou infraestruturas impróprias, tais como documentos físicos, *pen drives*, e-mail e computadores pessoais, acarretando a inatividade e perda desses dados (Silva, 2019). Essa prática é denominada de “fenômeno da gaveta de ficheiro” (Ferguson *et al.*, 2014, p. 1443), em que estudos revelam “[...] que mais de 50% das descobertas científicas não aparecem na literatura publicada [...]” (Ferguson *et al.*, 2014, p. 1443), sem citar a imensa quantidade de dados que são perdidos durante o processo da pesquisa científica pela falta de informação do pesquisador.

O reconhecimento do potencial informacional dos dados de pesquisa para a ciência contemporânea transforma a visão que os caracterizava como simples subprodutos dos processos de pesquisa. Naquele contexto, os dados eram considerados somente na sua configuração final, sem considerar os seus ciclos de vida, versões e linhagens e, via de regra, eram descartados ou armazenados em mídias ou em servidores sem a devida gestão quando os projetos eram concluídos. Quase sempre eram tragados silenciosamente

pelo tempo: pela obsolescência tecnológica e pela fragilidade das mídias digitais (Sayão, 2015, p. 5).

Nesse contexto, um insumo potencialmente valioso para o avanço da ciência era desprezado ou negligenciado pela falta de um adequado gerenciamento que envolve a coleta, tratamento, arquivamento, compartilhamento e preservação. Essas ocorrências produzem diversas implicações negativas para o ecossistema científico, como: impedimento do reuso dos dados, prejuízos para a visibilidade das produções científicas e a falta de transparência, credibilidade e integridade para a Ciência (Ball, 2014; Ferguson *et al.*, 2014; Sayão; Sales, 2015; Príncipe *et al.*, 2020; Vidal; Gabriel Júnior; Pavão, 2023).

Nessa conjuntura, comprehende-se que o fazer científico é um campo em constante mudança, e adequar-se ao surgimento das novas exigências para publicação das descobertas científicas é imprescindível tanto para a sobrevivência profissional do pesquisador quanto para o desenvolvimento da ciência (Shintaku; Ferreira Júnior, 2020). Diante desse novo contexto em que os dados possuem alta relevância para a ciência, é cada vez mais necessária a colaboração de órgãos institucionais no âmbito das IES para estabelecer uma ponte entre o pesquisador e as novas exigências do ecossistema científico relacionadas à Gestão de Dados Científicos (GDC). Assim, no cumprimento da sua função de difusora do conhecimento, é nesse contexto que as bibliotecas universitárias deverão estar inseridas.

Costa (2017, p. 87) afirma que “o surgimento do quarto paradigma traz impactos profundos sobre a ciência e, portanto, exige um profundo exame das funções das instituições empenhadas no avanço da ciência e no apoio aos cientistas, dentre elas as bibliotecas”. As bibliotecas universitárias são responsáveis por cumprir o papel de captar e difundir o conhecimento para as comunidades da universidade, com práticas que deverão estar alinhadas de forma contínua com as transformações tecnológicas e comunicacionais ocorridas na sociedade, ao mesmo tempo que respondem às demandas das necessidades institucionais de ordens política e social (Nunes; Carvalho, 2016).

Cunha (2010) ressalta que a adaptação das bibliotecas universitárias em uma sociedade cada vez mais conectada não deve estar focada unicamente em incorporar novas ferramentas tecnológicas em seu cotidiano, mas também, acima de tudo, em

compreender que essa adaptação passa pela reflexão de entender as reais necessidades da comunidade universitária, com o objetivo de promover serviços e produtos que atendam a essas necessidades. Nessa perspectiva, as bibliotecas universitárias são os órgãos que podem participar como protagonistas do processo de planejamento e oferta de serviços e produtos para a gestão dos dados científicos, juntamente a outras instâncias da IES.

Jones, Pryor e Whyte (2013) afirmam que o fornecimento de serviços para GDC exige múltiplos conhecimentos e deve ser planejado conjuntamente por diferentes instâncias da IES, como a alta gestão, programas de pós-graduação, profissionais da tecnologia da informação e pesquisadores. Porém, as bibliotecas, por se configurarem em um setor que já possui uma *expertise* no tratamento e organização da informação, podem liderar o planejamento de padrões, requisitos, suporte e soluções para este propósito.

Na concepção de Cox e Pinfield (2014), quando as bibliotecas se propõem a disponibilizar serviços de GDC, um dos pontos principais que devem ser levados em consideração é a heterogeneidade dos dados e as particularidades culturais existentes de cada área do conhecimento. Os autores explicam que esses elementos representam grandes desafios para as bibliotecas e, consequentemente, exigem serviços multifacetados e com abordagens holísticas.

Silva (2016) afirma que a oferta de serviços e produtos para GDC pelas bibliotecas deve se iniciar pela aplicação de uma entrevista. Este instrumento possibilita ao bibliotecário entender as reais necessidades dos pesquisadores para selecionar o melhor serviço para o contexto. Outra etapa importante é a identificação e elucidação de termos que são utilizados na biblioteconomia, mas que podem possuir significados díspares em outras áreas; esse monitoramento visa evitar ruídos na comunicação entre os bibliotecários e os pesquisadores.

Huang, Cox e Sbaffi (2021) corroboram com a afirmação de Silva (2016) ao relatar que as bibliotecas devem conhecer as necessidades dos pesquisadores antes de conceber os serviços para GDC, pois, sem essa medida, o serviço pode tornar-se subutilizado.

Nas palavras de Steeleworthy (2014, p. 2), os serviços e produtos para o gerenciamento de dados científicos podem ser definidos pela conjunção de:

[...] um sistema de pessoas, políticas, recursos e tecnologia que dão suporte e orientação a pesquisadores e organizações à medida que produzem, coletam, usam e preservam dados de pesquisa. Suas atividades e responsabilidades são abrangentes, um fato que pode afetar as maneiras pelas quais os profissionais de GDC enfatizam seus diferentes elementos (Steeleworthy, 2014, p. 2).

As práticas realizadas nas bibliotecas da Universiteit Leiden²⁸, University of Toronto²⁹ e The University of Edinburgh³⁰ para proporcionar suporte aos seus pesquisadores no gerenciamento dos seus conjuntos de dados encontram consonância com as afirmações de Steeleworthy (2014). Os Serviços e Produtos para o Gerenciamento de Dados Científicos (SPGDC) aplicados nessas bibliotecas fazem parte de práticas que abrangem toda a comunidade científica dessas universidades. O objetivo é inserir as recomendações da Ciência Aberta focada na abertura dos dados científicos no centro dos projetos desenvolvidos pelas bibliotecas para promover a pesquisa interdisciplinar, construções de novas ideias para fomentar a inovação e fortalecer a qualidade e a integridade dos resultados de pesquisa. Todos os projetos são coordenados por profissionais bibliotecários que são especialistas em gestão de dados científicos. No site das bibliotecas dessas universidades, podemos encontrar diversos SPGDC que são ofertados aos pesquisadores:

Consultoria para gerenciamento de dados – consiste em prestar orientações aos pesquisadores de forma individual ou em grupo nos seus procedimentos para gerenciar seus conjuntos de dados nas seguintes etapas: armazenamento, organização, documentação, arquivamento, *backup*, compartilhamento e publicação dos dados; criação do plano de gestão de dados; proteção de dados e orientação sobre o gerenciamento de dados pessoais e sensíveis; encontrar dados úteis para pesquisa; licenciamento de dados; encontrar repositórios de dados científicos que mais se adequem aos conjuntos de dados dos pesquisadores; preservação digital de dados e depósito em repositórios de dados científicos.

Capacitações em gestão de dados científicos – são oferecidos cursos, oficinas, palestras e *webinars* para capacitar os profissionais bibliotecários e a comunidade universitária no gerenciamento dos dados científicos, como, por exemplo:

²⁸ Ver informações em: <https://www.library.universiteitleiden.nl/researchers/data-management>

²⁹ Ver informações em: <https://oneresearch.library.utoronto.ca/researchdata/services-and-tools>

³⁰ Ver informações em: <https://library.ed.ac.uk/research-support/research-data-service/research-data-training-skills/workshops-courses>

tornar os dados em dados *FAIR*; construção do Plano de Gestão de Dados (PGD); publicação de dados; documentar dados e palestras e oficinas voltadas para os discentes de doutorado e áreas específicas do conhecimento.

Produtos ofertados – formulário de perguntas frequentes (FAQ) sobre o gerenciamento de dados; documentos modelos para construção do plano de gestão de dados; vídeos com informações sobre a gestão de dados científicos; catálogos com recursos de gerenciamento de dados científicos; documentos normativos para proceder com o gerenciamento de dados pessoais e sensíveis; políticas de gerenciamento de dados; guia de segurança de dados; *software* para construção do PGD; *software* para armazenamento de dados durante o desenvolvimento da pesquisa e repositórios de dados científicos.

Nessa perspectiva, as bibliotecas que aspiram incorporar tais serviços e produtos em seu catálogo, indiscutivelmente, deverão estabelecer planejamentos, políticas, conhecer as necessidades de sua comunidade, promover parcerias com outras instâncias da universidade e focar em um dos pontos mais fundamentais da oferta de SPGDC: a promoção da capacitação do seu corpo funcional para que possam se inserir nessa nova cultura centrada em dados que abrange o antes, o durante e o depois da pesquisa.

As três bibliotecas estrangeiras citadas, por exemplo, possuem profissionais bibliotecários capacitados em dados científicos para desempenhar funções que possam proporcionar para suas comunidades científicas uma variedade de SPGDC.

A formação do bibliotecário proporciona vasta compreensão sobre como organizar, tratar, armazenar, disseminar e recuperar conteúdos, ferramentas e funcionalidades que adicionam valor aos conteúdos digitais, e, com isso, o bibliotecário tem condições de colaborar no suporte e na gestão de dados (Sales *et al.*, 2019, p. 312).

Contudo, é necessário que o profissional da informação que não está inserido neste emergente universo de SPGDC reavalie e redimensione as suas habilidades para desempenhar esse novo papel como bibliotecário de dados.

[...] esses profissionais precisam ampliar suas competências para acompanhar as mudanças no âmbito da pesquisa científica, além de compreender o mundo complexo e heterogêneo dos dados de pesquisa, para que possam desenvolver serviços que atendam às necessidades dos pesquisadores em suas instituições, abrangendo as dimensões tecnológicas, legais e éticas (Sales *et al.*, 2019, p. 312).

Segundo Semeler e Pinto (2020), os bibliotecários de dados são profissionais que nascem das necessidades de organização do grande dilúvio de dados na pesquisa científica e da junção proveniente das áreas da Ciência da Informação, e-science e Ciência de Dados. Esses profissionais são capacitados para a criação de serviços e atendimento de demandas relacionadas ao gerenciamento de dados. Entretanto, o estudo desenvolvido por Fearon *et al.* (2013) com bibliotecas dos Estados Unidos que são vinculadas à *Association of Research Libraries (ARL)* revelou que apesar de os bibliotecários relatarem ter habilidades intrínsecas para prestar suporte na GDC, ainda assim admitem que é fundamental capacitar o corpo técnico das bibliotecas com novas habilidades e treinamentos para alcançar um serviço ideal de GDC.

Nesse seguimento, a pesquisa conduzida por Hu e Tang (2019) com 241 bibliotecários de 29 países corrobora com a pesquisa de Fearon *et al.* (2013), à medida que revela o desejo da maioria dos bibliotecários que compõe a amostra da pesquisa em receber mais treinamentos e cursos antes de expandir os serviços que já são prestados à comunidade científica. Esses profissionais também reivindicam um maior comprometimento institucional para a disponibilização de verbas para promover capacitações voltadas para bibliotecários.

Nas palavras de Tenopir *et al.* (2016), os serviços de GDC podem ser divididos em duas tipologias: informativos/consultivos, que são focados no desenvolvimento de guias, informações na web, políticas, orientações consultivas para a escolha de padrões de metadados e treinamentos; técnicos/práticos, com o objetivo de criar padrões de metadados, concepção de repositórios de dados científicos, curadoria de dados e seleção de conjuntos de dados. Pelo fato de os serviços informativos/consultivos exigirem menos recursos e um conhecimento menos aprofundado sobre a temática de dados, as bibliotecas podem começar a oferta serviços dessa tipologia.

No entendimento de Jones, Pryor e Whyte (2013), para garantir o sucesso na eficácia e eficiência dos serviços para GDC, é altamente recomendado que um planejamento anteceda sua concepção. A fase do planejamento visa garantir que o conjunto de serviços e produtos sejam desenvolvidos com objetivos lógicos e coerentes em suas finalidades e prazos. Nessa etapa, é imprescindível que seja realizada o reconhecimento das diferenças culturais na instituição, sensibilizar as

diferentes unidades da instituição para colaborar com o projeto, selecionar os responsáveis pela concepção de cada serviço e produto, levantamento do perfil do público, realizar uma análise dos requisitos necessários para criação dos serviços e produtos e o estabelecimento de um projeto-piloto.

O planejamento também passa pela criação de estratégias para capacitar os profissionais da biblioteca, o que envolve a identificação e compreensão de alguns elementos citados de forma recorrente por algumas instituições e autores como pré-requisitos para promover serviços e produtos para o gerenciamento dos dados científicos. Entre esses elementos, podemos citar: exigências das editoras científicas e agências de fomento à pesquisa para o compartilhamento de dados (PDG); classificação dos dados; ciclo de vida dos dados; Princípios FAIR; repositórios de dados e políticas para gestão de dados (Higgins, 2008; Martínez-Uribe; Fernández, 2015; Ruusalepp, 2008; Sayão; Sales, 2015; Silva, 2016 Plale; Kouper, 2017; Lima, 2019; Semeler; Pinto, 2019; Silva, 2019; Shintaku; Ferreira Junior, 2020; Príncipe *et al.*, 2020; Unesco, 2022; Vidal; Gabriel Júnior; Pavão, 2023).

Os elementos citados anteriormente possuem interligações e uma forte influência entre si no contexto da Ciência Aberta. A observação desses elementos influencia a oferta de serviços e produtos elencados a partir da literatura nacional e internacional, conforme apresentados no Apêndice C. Nas próximas subseções, alguns desses elementos serão abordados de forma separada com o objetivo de proporcionar um aprofundamento de informações acerca de suas exigências, características, conceitos e alcance.

5.1 Editoras científicas

Nos últimos anos, algumas editoras científicas (*publishers*), como a *PLOS One*, *Springer Nature*, *Elsevier* e *Taylor & Francis*, passaram a incorporar em suas políticas editoriais diversas orientações e exigências para que os pesquisadores disponibilizassem os seus dados científicos. As políticas contêm informações de como funciona o processo de abertura dos dados científicos dos pesquisadores que pretendem publicar em seus periódicos. A seguir, serão abordados alguns pontos contemplados nessas políticas editoriais de forma resumida (Shintaku; Ferreira Junior, 2020).

PLOS One – possui assinatura de declaração de conformidade com as políticas da revista, disponibilização mínima de um conjunto de dados (dados brutos); depósito em um repositório com identificador persistente (DOI) e um conjunto de metadados que possam descrever os dados para o seu acesso e localização; permissão para não compartilhamento de dados que envolvam questões éticas, jurídicas e de direitos autorais, como dados de terceiros, dados de participantes de pesquisa, dados sensíveis e dados privativos, sendo que essas restrições para os dados não contemplam razões de ordem pessoal dos pesquisadores.

Springer Nature – a editora decidiu utilizar modelos de níveis de abertura para a disponibilização dos dados, sendo que o primeiro nível é o mais flexível e recomendado para periódicos que estão começando no processo da abertura dos dados e o último nível, mais rígido, recomendado para periódicos mais experientes neste processo. Os periódicos que fazem parte desse grupo editorial podem optar por quais níveis de abertura serão adotados em suas políticas editoriais, existindo a possibilidade que esses periódicos, caso necessitem, ampliem essas exigências e recomendações. Os modelos são divididos em quatro: a) o compartilhamento de dados e citação dos dados é encorajado; b) compartilhamento de dados e dados de evidência é encorajado; c) compartilhamento de dados é incentivado e a declaração de disponibilidade de dados é necessária; d) compartilhamento de dados, dados de evidência e dados de revisão são necessários.

Elsevier – a política da editora, diferentemente dos outros *publishers*, não possui uma diretriz mandatória, e sim um viés pautado para apontar recomendações e incentivos na disponibilização dos dados. A política ainda chama a atenção para que, durante o processo de abertura dos dados, algumas questões devem ser levadas em consideração, tais como: a multiplicidade tipológica dos dados por conta das diferenças disciplinares (isso implica um tratamento mais complexo dos dados) e a curadoria para proporcionar a integridade dos dados

Taylor & Francis – a política apresenta níveis com diferentes graus de abertura para os dados. Os níveis de abertura podem ser escolhidos pelos mais de 2000 periódicos que fazem parte deste grupo editorial, seguindo uma lógica semelhante à da política da Springer Nature. Os níveis de abertura estão divididos em cinco: a) incentiva a disponibilização ou compartilhamento dos dados; b) disponibilização ou compartilhamento dos dados mediante uma solicitação razoável, a escolha pelo

compartilhamento fica a critério do autor; c) a disponibilização ou compartilhamento é obrigatória, o autor escolhe a licença de uso que envolve este processo; d) a disponibilização ou compartilhamento é obrigatória e o autor deverá optar por licenças de uso que permitam a utilização dos seus dados de forma irrestrita, os dados também devem ser localizáveis e acessíveis; e) a disponibilização ou compartilhamento seguem as mesmas regras dispostas no item d), além de atender às iniciativas dos Princípios FAIR.

As recomendações da Ciência Aberta para periódicos científicos não se restringem apenas às solicitações e exigências para os pesquisadores compartilharem os seus dados científicos, mas adota outras diversas práticas como o estabelecimento do fluxo contínuo de publicações, critérios de aceitação de pré-prints, revisão por pares aberta e o uso do *Digital Object Identifier*³¹ (DOI), *Open Researcher and Contributor ID*³² (ORCID) e o *Research Organization Registry*³³ (ROR).

No Brasil, podemos citar como exemplo de adoção de práticas para a abertura dos dados científicos a *Journal of Venomous Animals and Toxins Including Tropical diseases*, a *Memórias do Instituto Osvaldo Cruz* e a *Revista da Sociedade Brasileira de Medicina Tropical*. As três revistas participaram de um projeto piloto da SciELO para promover a adoção das recomendações da Ciência Aberta pelas revistas indexadas em sua base de dados. Entre as recomendações adotadas pelas revistas, destacamos o requerimento para o compartilhamento dos dados brutos das pesquisas como condicionante para publicação em seus periódicos (Barra Vieira *et al.*, 2021).

Nas revistas brasileiras na área de Ciências Sociais Aplicadas indexadas na SciELO, uma pesquisa realizada por Dias (2024) revelou que apenas 34% possuem algum tipo de política para compartilhamento de dados, 87,2% não possuem política mandatária para o compartilhamento de dados, 91,5% não possuem adesão aos Princípios FAIR e apenas uma faz recomendações para uso de um buscador persistente. Nesta conjuntura, a pesquisa revela um grande prejuízo para o ecossistema da ciência na medida em que as práticas citadas são consideradas fundamentais para possibilitar o acesso aos conjuntos de dados das pesquisas científicas para a sua reproduzibilidade e o seu reuso.

³¹ Ver em: <https://www.doi.org/>.

³² Ver informações em: <https://info.orcid.org/pt/what-is-orcid/>.

³³ Ver informações em: <https://ror.org/>.

A recomendação da Ciência Aberta que possui maior adesão entre os periódicos científicos brasileiros é a disponibilização de seus artigos em acesso aberto. Um estudo realizado por Pereira e Furnival (2020) constatou que o país ocupa a 3^a colocação mundial na adesão dessa modalidade, com 1458 periódicos em acesso aberto. Porém, quando a adesão é relacionada a outras recomendações da Ciência Aberta, o processo se encontra em vias de desenvolvimento, com baixa adesão dos periódicos científicos, principalmente quando o critério envolve exigências para o compartilhamento de dados científicos.

5.2 Políticas para o compartilhamento de dados científicos

As políticas públicas são um dos fatores preponderantes no estabelecimento de diretrizes para alcançar o desenvolvimento de um país, ao mesmo tempo que visam solucionar um problema público identificado. A construção de políticas públicas, visa atender também às reivindicações de grupos da sociedade que convergem e divergem em seus objetivos nessa estrutura social. Nessa perspectiva, Bonetti (2010, p. 74) define as políticas públicas como “o resultado da dinâmica de jogo de forças que se estabelece no âmbito das relações de poder, relações essas constituídas pelos grupos econômicos e políticos, classes sociais e demais organizações da sociedade civil”.

Nesse sentido, é fundamental que a comunidade científica, como participante dessa estrutura social, entenda como funcionam essas relações de poder existentes. Isso permitirá que ela se organize para criar estratégias e caminhos para influenciar a formulação de políticas abertas em prol da ciência. Conforme Costa (2019), “As políticas abertas são documentos normativos de países, instituições de pesquisa e agências de fomento que promovem as práticas da ciência aberta”.

Como consequência da organização da comunidade científica, políticas públicas foram desenvolvidas em diversos países com objetivo de criar diretrizes para que os dados pudessem ser gerenciados adequadamente e posteriormente compartilhados. Uma grande parte dessas políticas foi criada a partir das recomendações da OCDE de 2007 (*Principles and Guidelines for Access to Research Data from Public Funding*), que incentiva o compartilhamento dos dados das pesquisas financiadas com verba pública aos países membros do organismo.

As Recomendações da OCDE estabelecem normas ou objetivos coletivos e precisos que se espera que os países membros implementem. Uma “Recomendação” é um instrumento jurídico da OCDE que não é juridicamente vinculativo, mas que, através de uma prática de longa data dos países membros, é considerada como tendo uma grande força moral. As recomendações da OCDE são adotadas quando os governos membros estão preparados para assumir um compromisso político para implementar os princípios (e/ou diretrizes) nelas estabelecidos (OCDE, 2007, p. 7).

No estudo desenvolvido por Ruusalep (2008), com apoio da *Data Digital Curation*, foi realizado um levantamento de algumas políticas criadas através das recomendações da OCDE (2007). Com base nesse estudo, serão relatadas iniciativas realizadas em alguns países para a criação de políticas para o compartilhamento de dados científicos, publicadas por governos nacionais, agências de fomento à pesquisa e institutos de ensino e pesquisa. Além disso, foram incluídas fontes de informações mais atuais para complementar a exposição.

No Reino Unido, a *Research Councils United Kingdom (RCUK)* elaborou, em 2006, uma declaração intitulada *Position Statement on Access to Research Outputs*, que se tornou um documento referencial para que as instituições que fazem parte da sua tutela elaborassem as suas próprias declarações para gestão de dados, com, as seguintes orientações:

- a) As ideias e conhecimentos derivados de investigação financiada com fundos públicos devem ser disponibilizados e acessíveis para uso público, interrogatório e escrutínio, de forma tão ampla, rápida e eficaz quanto possível; b) Os resultados da investigação publicados devem estar sujeitos a uma rigorosa garantia de qualidade, através de uma revisão por pares com mecanismos eficazes; c) Os modelos e mecanismos de publicação e acesso aos resultados da investigação devem ser eficientes e economicamente acessíveis na utilização de fundos públicos; d) Os resultados da investigação atual e futura devem ser preservados e permanecer acessíveis para futuras gerações (Ruusalep, 2008).

Já em 2011, o *RCUK* elaborou a *Common Principles on Data Policy* com princípios obrigatórios a serem seguidos pelos conselhos de pesquisa do Reino Unido. No entanto, a *RCUK* permitiu que adaptações pudessem ser realizadas nos princípios para se enquadrar no contexto político de cada conselho de pesquisa (Santos *et al.* 2017). Atualmente a *RCUK* é reconhecida pelo nome de *United Kingdom Research and Innovation (UKRI)*. Além da *UKRI*, o Reino Unido ainda possui outras instituições que financiam pesquisas científicas: *National Institute for Health and Care Research*

(NIHR)³⁴, Cancer Research UK (CRUK)³⁵, Wellcome Trust³⁶ e a British Heart Foundation UK³⁷. Todas as instituições citadas possuem diretrizes para que os pesquisadores que sejam financiados com as suas verbas compartilhem os seus dados (Ruusalep, 2008).

Em 2014, foi implementado o *Research Excellence Framework (REF)*, uma política contendo vinte diretrizes com a finalidade de avaliar as instituições do Reino Unido no que tange à qualidade das suas pesquisas científicas. Entre as diretrizes, a REF recomenda uma “[...] maior transparência e abertura nas infraestruturas de dados” (Santos *et al.*, 2017, p. 66) das instituições do Reino Unido.

Em 2016, o Reino Unido formulou uma política para gerenciamento de dados científicos intitulado *Concodat on Open Research Data*. A política possui dez princípios a serem seguidos por instituições britânicas para o compartilhamento de dados. Sua elaboração contou com a participação de quatro instituições, *Higher Education Funding Council for England (HEFCE)*, *Research Council UK*, *Universities UK* e *Wellcome Trust*, além de sugestões da comunidade científica (Santos *et al.* 2017).

Nos Estados Unidos, a *National Science Fundation (NSF)* foi uma das primeiras agências de fomento à pesquisa a formular políticas com recomendações para o compartilhamento dos dados científicos das pesquisas financiadas com verba pública (Atkins *et al.*, 2003). Neste sentido, suas divisões e programas passaram a criar de forma específica as suas próprias políticas. No entanto, algumas divisões e programas, diferentemente da NSF, criaram diretrizes mandatórias para o compartilhamento dos dados. Seguindo regras semelhantes da NSF, outras organizações, tais como a *American Sociological Association*, *American Psychological Association*, *National Institutes of Health (NIH)* e *National Institutes of Justice*, também passaram a exigir o compartilhamento de dados das pesquisas aos contemplados por seus financiamentos (Ruusalep, 2008).

Em 2013, foi criada a *Increasing Access to the Results of Federally Funded Scientific Research*, um memorando emitido pelo governo federal dos EUA como

³⁴ Ver informações em: <https://nhr.opendatasoft.com/pages/homepage/>

³⁵ Ver informações em: <https://www.cancerresearchuk.org/funding-for-researchers/applying-for-funding/policies-that-affect-your-grant/data-sharing-and-management-policy>

³⁶ Ver informações em: <https://wellcome.org/what-we-do/our-work/research-enviorment/open-research>

³⁷ Ver informações em: <https://bhfdata-science-centre.org/publication-and-dissemination-policy/>

forma de incentivar a criação de políticas pelas agências de fomento à pesquisa e institutos de ensino e pesquisa do país. As políticas criadas deveriam conter regras objetivas para permitir o acesso aos dados científicos das pesquisas oriundas de financiamento público, além de prever a participação de pesquisadores, universidades, bibliotecas, editoras e sociedade civil na formulação dessas políticas (Santos *et al.*, 2017).

Ainda em 2013, foi formulada a *Open Data Policy - Managing Information as an Asset (M-13-13)* para estabelecer o gerenciamento de dados de diversas instituições nos Estados Unidos. Já em 2014, como forma de aplicar as recomendações relacionadas ao gerenciamento de dados do M-13-13, é criado o *Open Data Plan*, uma política nacional para facilitar o acesso da população aos dados do governo, inclusive os dados das instituições de pesquisa e ensino (Santos *et al.*, 2017).

Na Austrália, em 2009, foi fundada a *Australian National Data Service (ANDS)*, um órgão do governo australiano responsável em promover a reunião dos dados científicos das instituições de pesquisa do país a partir de uma infraestrutura eletrônica para facilitar o acesso aos dados de diversas instituições do país. A ANDS também foi responsável pela criação de guias para gestão de dados voltados para pesquisadores e instituições (Santos *et al.*, 2017).

Posteriormente, com a criação de uma política nacional intitulada *Accessibility Framework*, a Austrália passa a contar com uma série de diretrizes para que os dados científicos oriundos de financiamentos públicos fossem compartilhados, acessíveis e detectáveis, através do estabelecimento de repositórios para este propósito. As diretrizes criadas no *Accessibility Framework* foram utilizadas como referencial para que outros organismos de fomento à pesquisa na Austrália, como *The Australian Research Council (ARC)* e a *National Health and Medical Research Council (NHMRC)*, pudessem atualizar as suas políticas para obrigar o compartilhamento dos dados das pesquisas financiadas pelos organismos. (Ruusalep, 2008)

Já em 2015, o governo australiano elabora a *Australian Government Public Data Policy Statement*, uma política pública nacional para normatizar o compartilhamento de dados gerados por diversos órgãos australianos, inclusive as instituições de pesquisa e ensino. A Declaração de Política informa que todos os dados não sensíveis gerados pelo governo passam a ser abertos por padrão. Para

garantir o acesso a esses dados, uma infraestrutura eletrônica foi disponibilizada para garantir o serviço. (*Government Australian*, 2015; Santos *et al.*, 2017)

Em 2023, o governo australiano lança o programa *Underpinnings* através da coordenação da *Australian Research Data Commons* (ARDC), com a colaboração de vinte e cinco universidades. O Programa coletou informações durante os anos de 2021 e 2022 com a proposta de identificar elementos essenciais para o gerenciamento de dados científicos em universidades e institutos de pesquisa. Ao término da coleta, foram identificados dezenove elementos, sendo que nove foram considerados essenciais para um gerenciamento de dados científicos com necessidades de aplicação imediata. Entre os nove elementos essenciais em gerenciamento de dados científicos, destaca-se o elemento três, que aborda a necessidade de construção de políticas para balizar as decisões que serão tomadas no que tange aos dados científicos (ARDC, 2023).

Uma política de gerenciamento de dados científico eficaz oferece a instituição uma abordagem estruturada para atender aos seus requisitos regulatórios e garantir que as funções, responsabilidades, processos e procedimentos necessários estejam em vigor para um gerenciamento de dados científicos eficaz (ARDC, 2023, p. 2).

Neste aspecto, a ARDC e as universidades que participaram do projeto entendem que a construção de políticas para o gerenciamento de dados científicos é fundamental para planejar todas as tomadas de decisões referentes a dados científicos, como: mudança cultural, construção de infraestruturas, retenção e descarte, publicação, treinamentos, preservação digital e financiamento.

No Canadá, Alemanha, Finlândia, Espanha, Holanda e Portugal, também foram criadas políticas nacionais voltadas para o incentivo ao compartilhamento dos dados de pesquisas financiadas com verba pública. A intenção era criar um documento de referência e apresentar direcionamentos aos organismos de fomento à pesquisa e as instituições de ensino e pesquisa destes países no tocante ao desenvolvimento de suas próprias políticas (Ruusalep, 2008; Santos *et al.*, 2017).

No Brasil, as exigências para o compartilhamento dos dados das pesquisas financiadas com verba pública pelas agências de fomento à pesquisa ainda são bastante incipientes. As principais instituições envolvidas no financiamento à pesquisa no país são: a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes), o Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq),

a Financiadora de Estudos e Projetos (Finep) e o Conselho Nacional das Fundações Estaduais de Amparo à Pesquisa (CONFAP), composto por 27 instituições, cada uma localizada em uma unidade federativa. Porém, dessas instituições, apenas a Fundação de Amparo à Pesquisa de São Paulo (FAPESP) possui exigências para a construção do Plano de Gestão de Dados (PGD) como requisito para concessão de suas bolsas de financiamento, sendo que as exigências para a construção do PGD estão vigentes desde 2017³⁸.

Entre as políticas para o gerenciamento de dados científicos por instituições de ensino e pesquisa no Brasil, destacam-se as políticas da Fiocruz, Embrapa e IBICT, intituladas respectivamente como: Política de Gestão, Compartilhamento e Abertura de Dados para Pesquisa e Política de Governança de Dados, Informação e Conhecimento³⁹ e Plano de Dados Abertos 2018-2019⁴⁰. As políticas possuem princípios e diretrizes detalhadas para o estabelecimento da coleta, tratamento, preservação e compartilhamento dos dados científicos das instituições.

O início para a criação de políticas mais elaboradas para o desenvolvimento da Ciência Aberta e o compartilhamento de dados no Brasil surge devido à participação do país na *Open Government Partnership – OGP* (Parceria para Governo Aberto). A OGP é uma organização multilateral criada em 2011, a entidade possui 75 países membros e tem o Brasil como um de seus cofundadores. Os países que fazem parte da OGP se comprometem a desenvolver planos de ação nacionais a cada dois anos que versem sobre a criação de políticas, ações e infraestruturas para a resolução de problemas públicos em áreas de combate à corrupção, acesso à informação, fortalecimento da democracia, participação social, desenvolvimento da ciência e o compartilhamento de dados (Parceria para o Governo Aberto, 2018). Neste contexto, a Parceria para o Governo Aberto no Brasil, nos últimos anos, estabeleceu a criação de diversas ações para o desenvolvimento da Ciência Aberta no país, que serão descritas a seguir.

Em 2016, é criado o Decreto nº 8.777, que institui a Política de Dados Abertos do Poder Executivo Federal. Destaca-se entre os objetivos da Política, “promover a

³⁸ Ver informações em: <https://www.abcd.usp.br/noticias/fapesp-comeca-exigir-plano-de-gestao-de-dados/>

³⁹ Ver informações em:

<https://www.embrapa.br/documents/10180/1600893/Pol%C3%ADtica+de+Governan%C3%A7a+de+Dados%2C+Informa%C3%A7%C3%A3o+e+Conhecimento+da+Embrapa/96873780-8e04-b67d-be7c-a0b6a4318556>

⁴⁰ Ver informações em: <http://sitehistorico.ibict.br/servico-de-informacao-ao-cidadao-1/PlanodeDadosAbertos20182019.pdf>

publicação de dados contidos em bases de dados de órgãos e entidades da administração pública federal direta, autárquica e fundacional sob a forma de dados abertos" (Brasil, 2016) e a construção do Plano de Dados Abertos (PDA) pelos órgãos que fazem parte do Poder Executivo Federal.

O PDA é uma política que deve ser publicada e reformulada a cada dois anos. Seu conteúdo é composto por princípios e diretrizes de como será gerido os dados produzidos pelas instituições. Sua criação visa atender também a Lei de Acesso à Informação (Lei nº 12.527⁴¹) e, desde 2018, os PDAs devem ser construídos em conformidade com a Lei Geral de Proteção de Dados (Lei nº 13.709/2018⁴²).

Plano de Dados Abertos é o documento orientador para as ações de implementação e promoção de abertura de dados, obedecendo a padrões mínimos de qualidade e acessibilidade, de modo a facilitar o entendimento e a reutilização de informações (Controladoria Geral da União, 2025).

A normatização da formulação dos PDAs é regida pela Resolução nº 3, de 13 de outubro de 2017⁴³, e fica sob a incumbência da Controladoria Geral da União (CGU) a fiscalização da formulação dos PDA. O cumprimento da formulação dos PDAs pelos órgãos da administração direta, autárquica e fundacional pode ser acessado através do Painel de Monitoramento de Dados Abertos⁴⁴, no qual podem ser conferidos os PDAs atrasados, em andamento e os publicados (Brasil, 2016). Segundo os dados do Painel de Monitoramento de Dados Abertos, até o ano de 2024, das duzentas e sessenta e sete instituições que devem formular seu PDA, cento e cinquenta e cinco estão com a formulação da política em atraso. Entre as 14 (quatorze) universidades que compõem a amostra dessa pesquisa, nove estão com o PDA em atraso (Controladoria Geral da União, 2024). Isso causa grande preocupação, pois é justamente através dessas políticas que são criadas diretrizes e princípios para proceder com o gerenciamento de dados científicos dessas instituições.

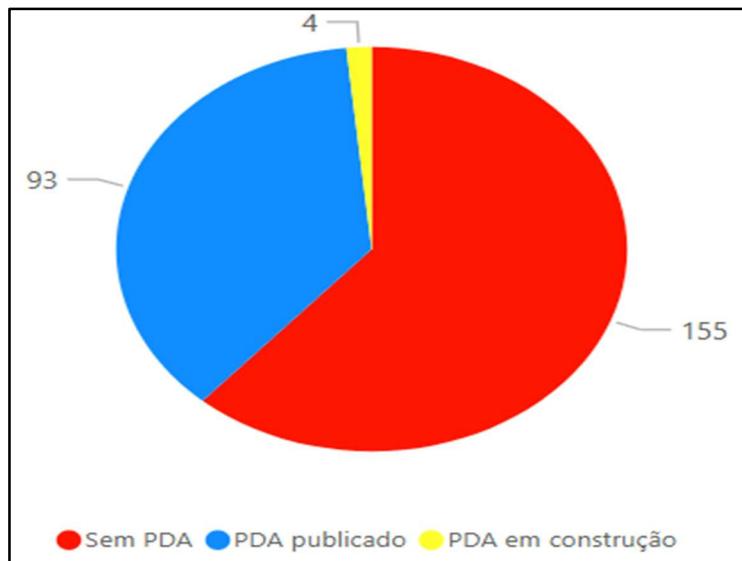
⁴¹ Ver informações em: https://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/l12527.htm

⁴² Ver informações em: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm

⁴³ Ver informações em: <https://www.gov.br/conard/pt-br/legislacao-arquivistica/resolucoes/resolucao-n-o-3-de-13-de-outubro-de-2017>

⁴⁴ Ver informações em: <https://centralpaineis.cgu.gov.br/visualizar/dadosabertos>

Figura 4 – Status dos Planos de Dados Abertos



Fonte: Painel Monitoramento de Dados Abertos.

Em relação aos Planos de Ação Nacional em Governo Aberto, o Brasil já formulou seis Planos: 1º Plano (2011-2013⁴⁵); 2º Plano (2014-2016⁴⁶); 3º Plano (2016-2018⁴⁷); 4º Plano (2018-2021⁴⁸); 5º Plano (2021-2023⁴⁹) e o 6º Plano (2023-2027⁵⁰). Este último, diferentemente dos anteriores, teve sua vigência estabelecida para quatro anos.

É importante ressaltar que foi no 4º Plano, por meio do Compromisso 3 (Inovação e Governo Aberto na Ciência), que metas para governança de dados científicos foram estabelecidas pela primeira vez para o avanço da Ciência Aberta no Brasil. O Compromisso 3 foi formado por 9 marcos, sob a coordenação da Embrapa e com a participação do IBICT, Capes, CNPq, Fiocruz, OKBR, Associação Nacional de Pesquisa e Pós-Graduação em Ciência da Informação/UnB, RNP e a CNEN (Comissão Nacional de Energia Nuclear). As propostas e os resultados de cada marco do Plano estão expostos no Quadro 8 (Parceria para Governo Aberto, 2018).

⁴⁵ Ver informações em: <https://repositorio.cgu.gov.br/handle/1/66727>

⁴⁶ Ver informações em: <https://repositorio.cgu.gov.br/handle/1/66738>

⁴⁷ Ver informações em: <https://repositorio.cgu.gov.br/handle/1/66732>

⁴⁸ Ver informações em: <https://repositorio.cgu.gov.br/handle/1/66740>

⁴⁹ Ver informações em: <https://repositorio.cgu.gov.br/handle/1/67649>

⁵⁰ Ver informações em: https://www.gov.br/cgu/pt-br/governo-aberto/a-ogp/planos-de-acao/6deg-plano-de-acao-brasileiro/brazil_action-plan_2023-2027_december_pt.pdf

Quadro 8 – 4º Plano de Ação Nacional para o Governo Aberto

Compromisso 3. Estabelecer mecanismos de governança de dados científicos para o avanço da ciência aberta no Brasil	
PROPOSTAS	RESULTADOS
1 - Implantação de uma rede interinstitucional pela Ciência Aberta.	Criação da <i>Research Data Alliance Brazil</i> .
2 - Realização de diagnóstico nacional e internacional da Ciência Aberta	Livro - Panorama Preliminar do Desenvolvimento da Ciência Aberta.
3 - Definição de diretrizes e princípios para políticas institucionais de apoio à Ciência Aberta	Manual - Princípios e Diretrizes Orientadoras para Políticas Institucionais de Dados Científicos Abertos em Apoio a Ciência Aberta.
4 - Promoção de ações de sensibilização, participação e capacitação em Ciência Aberta	1 - Criação de vídeos: A experiência da Ciência Aberta na pesquisa científica; Interoperabilidade e infraestrutura tecnológica na Ciência Aberta; A Ciência Aberta na perspectiva das agências de fomento; A Ciência Aberta na perspectiva dos editores científicos; Gestão de Dados e Princípios FAIR; O Movimento da Ciência Aberta; Webinar - O papel da Ciência Aberta e Governo Aberto face à pandemia do novo coronavírus covid-19. 2 - Curso EAD sobre Ciência Aberta, promovido pela Fiocruz ⁵¹ . 3 - Cartilha de Gestão de Dados de Pesquisa para Pesquisadores ⁵² .
5 - Articulação com agências de fomento para a implantação de ações de apoio à Ciência Aberta	1 - Aquisição de identificadores digitais persistentes a serem atribuídos a conjuntos de dados de pesquisa. 2 - Formação do Consórcio Nacional para Ciência Aberta, (CoNCienciA). 3 - Assinatura de acordo de cooperação para desenvolvimento e implantação do repositório de dados de pesquisa 'Lattes Data', pelo CNPq.
6 - Articulação com editores científicos para a implantação de	1 - Ações de sensibilização e comunicação em eventos técnico-científicos.

⁵¹ Ver informações em: <https://campusvirtual.fiocruz.br/portal/?q=formacao-modular/ciencia-aberta>⁵² Ver informações em:<https://livroaberto.ibict.br/bitstream/123456789/1083/2/cartilha%20dados%20de%20pesquisa.pdf>

ações em apoio à Ciência Aberta	<p>2 - Publicação dos livros digitais: Ciência Aberta para editores científicos⁵³ e Tópicos sobre dados abertos para editores científicos⁵⁴.</p> <p>3 - Implantação de dois repositórios de preprints: <i>SciELO Preprints</i>⁵⁵ e o <i>EmeRI (Emerging Research Information)</i>⁵⁶.</p>
7 - Implantação de infraestrutura federada piloto de repositórios de dados de pesquisa.	1 - Implantação de dois repositórios de dados científicos: Arca Dados da Fiocruz ⁵⁷ e o Deposita Dados do IBICT ⁵⁸ .
8 - Proposição de padrões de interoperabilidade para repositórios de dados de pesquisa.	Criação do Manual do Dataverse: contendo um conjunto mínimo de padrões de metadados para garantir a integração dos repositórios de dados de pesquisa, de forma que esses possam ser facilmente localizáveis, acessíveis, interoperáveis e reutilizáveis em nível nacional, regional e internacional, além de possuir orientações para os bibliotecários de como gerir a plataforma.
9 - Proposição de conjunto de indicadores para aferição da maturidade em Ciência Aberta	A construção de um conjunto de indicadores para aferir o grau de maturidade das instituições de C&T para a abertura de dados de pesquisa, organizados nas seguintes categorias: governança, cultura organizacional, gestão de dados de pesquisa e infraestrutura tecnológica.

Fonte: elaboração própria com base no 4º Plano de Ação Nacional em Governo Aberto e no Relatório Final de Autoavaliação do 4º Plano de Ação Nacional em Governo Aberto.

A continuidade do desenvolvimento da Ciência Aberta no Brasil foi seguida pela criação do 5º e 6º Planos de Ação Nacional em Governo Aberto. No 5º Plano, por meio do Compromisso 8, abordou-se o tema “Transparência em Ciência: novos mecanismos de avaliação para o avanço da Ciência Aberta”. As instituições participantes da sua construção objetivaram trabalhar para a resolução dos seguintes problemas públicos identificados: carência de diretrizes nas políticas das agências de fomento para o reconhecimento de práticas da Ciência Aberta em pesquisas científicas; carência de métricas que valorizem as práticas de Ciência Aberta e baixa

⁵³ Ver informações em:

https://www.abecbrasil.org.br/arquivos/Ciencia_aberta_editores_cientificos_Ebook.pdf

⁵⁴ Ver informações em:

https://www.abecbrasil.org.br/arquivos/Topicos_dados_abertos_editores_cientificos.pdf

⁵⁵ Ver informações em: <https://preprints.scielo.org/index.php/scielo>

⁵⁶ Ver informações em: <https://preprints.ibict.br/>

⁵⁷ Ver informações em: <https://arcadados-beta.fiocruz.br/>

⁵⁸ Ver informações em: <https://depositadados.ibict.br/>

conexão entre as plataformas voltadas para a ciência (Parceria para o Governo Aberto, 2021).

Já no 6º Plano, por meio do Compromisso 3, o tema escolhido foi “Práticas colaborativas para a ciência e a tecnologia”. O Plano, que está em vias de aplicação, destaca principalmente a necessidade da concepção de uma política para Ciência Aberta em âmbito nacional e problematiza em torno dos desafios relacionados ao compartilhamento de dados no Brasil, expondo as seguintes questões:

Qual o limite e as regras para a reutilização de dados científicos? Qual o prazo de embargo dos dados de pesquisa que culminou com o desenvolvimento tecnológico? O compartilhamento de dados pode ocorrer apenas com pesquisadores brasileiros (soberania nacional), ou pode ocorrer com pesquisadores estrangeiros? Se o pesquisador receber fomento de agências internacionais e elas exigirem o compartilhamento de dados como contrapartida do apoio, qual a orientação política do Brasil quanto ao caso? (Parceria para o Governo Aberto, 2023, p. 22).

A construção de políticas que exigem o compartilhamento dos dados científicos, principalmente em nível nacional, é uma prática que se alinha com as necessidades de desenvolvimento da ciência e da sociedade. Isso ocorre porque, através do acesso aos conjuntos de dados, múltiplas oportunidades são geradas no campo científico para transpor desafios sociais, além de oferecer uma resposta aos cidadãos que financiam a ciência com o pagamento de seus impostos. Nesse seguimento, no âmbito das IES, existe a necessidade de profissionais em biblioteconomia munidos de informações sobre as práticas em Ciência Aberta para influenciar na criação de políticas voltadas para o gerenciamento de dados científicos em suas instituições.

5.3 Classificações dos dados

Os diferentes formatos que os dados científicos possuem são uma característica que demonstra as suas múltiplas possibilidades para o desenvolvimento de outras pesquisas. No entanto, a sua heterogeneidade também poderá ser um dos principais obstáculos para o seu tratamento, caso este fato não seja levado em consideração. No entendimento de Sayão e Sales (2015), com a intenção de evitar a ocorrência dessa problemática, é fundamental que os responsáveis pelo suporte em

seu gerenciamento criem um planejamento que reconheça as variedades existentes dos dados para a adoção de critérios em seu adequado tratamento.

Nesta perspectiva, organismos e autores criaram alguns critérios para categorizar esses dados. Por exemplo, segundo Silva (2019, p. 29), em uma análise inicial, podemos encontrar uma distinção básica de dados entre quantitativos e qualitativos. Nesta interpretação, “[...] os primeiros são qualificáveis, ou seja, podem ser expressos numericamente. Os qualitativos, por outro lado, são de caráter descriptivo e tentam refletir a qualidade dos atributos dos objetos ou sujeitos”.

Existem outras análises em que os dados são classificados em diversas categorias com base nos procedimentos aplicados durante a sua constituição. O Quadro 9 a seguir demonstra um compilado de diversas classificações dos dados científicos (National Science Foundation, 2005; Sayão; Sales, 2015; Silva, 2019).

Quadro 9 – Classificação dos dados

Dados segundo a sua origem ou coleta	Observacionais:	São dados que têm caráter único e muito difícil de serem reproduzidos, por conta dos instrumentos utilizados e do contexto, por este motivo possuem um valor inestimado para ciência e geralmente são preservados por um período indeterminado. Exemplos: observações climatológicas, fotografias de uma supernova, levantamento de altitudes de uma comunidade, radares e sensores.
	Computacionais:	São dados provenientes de modelos de computador ou de ordenações e simulações computacionais ou de fluxos de trabalhos. Possuem preservação por tempo definido, pois são reproduzíveis, caso sejam documentadas todas as informações sobre os processos, fluxos, <i>hardware</i> e <i>software</i> pelos quais foram necessários para a sua concepção. Necessitam também de um modelo de metadados robusto para serem inseridos em um repositório. Exemplos: estudo das marés, interações políticas e o comportamento do mercado financeiro.
	Experimentais:	São dados provenientes de experimentos científicos, assim, como os computacionais possuem preservação por um tempo definido e necessitam também de uma documentação rigorosa de cada passo que foi executado para sua concepção, com o

		objetivo de permitir a sua reprodução. Exemplos: medições de padrões de expressão genética e taxas de reações químicas.
Dados segundo a sua natureza	Números, metodologias, protocolos, imagens, <i>software</i> , avaliações, planilhas, algoritmos, textos, banco de dados, notas de campo, sequências genéticas, coordenadas geográficas, equações, animações ou modelos e simulações.	
Dados segundo a sua fase de pesquisa	Primários, brutos, crus ou preliminares:	São dados que vêm diretamente dos instrumentos científicos, trabalhos originais de pesquisa e/ou dados brutos sem interpretação.
	Secundários ou Derivados:	São dados resultantes do processamento, da análise ou da combinação de dados brutos ou outros dados.
	Terciários, Canônicos ou Referenciais:	São coleções de dados consolidados e arquivados geralmente em grandes centros de dados
Dados segundo o grau de estruturação	Estruturados	São dados que possuem uma organização a ponto de permitir facilmente a sua transferência para outros sistemas, pois a sua organização segue um modelo de tabela que representa associações entre um termo e sua relação hierárquica.
	Semiestruturados	São dados que possuem um determinado grau de organização, como imagens e documentos de texto, podem ser facilmente transferidos para outros sistemas, entretanto, os seus dados não formam parte de uma base de dados, pois não possuem um esquema predefinido, assim, não mantém uma estrutura de base relacional.
	Não estruturados	Não têm um modelo de dado definido ou estrutura identificável, difíceis de ser combinados com outros dados ou de serem analisados. Geralmente os dados não estruturados são os qualitativos.
Dados Abertos	“São dados que podem ser utilizados e distribuídos por qualquer pessoa, sem barreiras técnicas ou legais, sob os requerimentos de reconhecer a autoria e compartilhar o novo produto nas mesmas condições (Silva, 2019, p. 33)”.	

Dados pessoais e sensíveis	"Identificadores como nomes ou números de identificação, características físicas, fisiológicas, genéticas, mentais, econômicas, culturais ou sociais, incluindo também dados de localização de GPS ou de telemóveis (OpenAIRE, 2024)
----------------------------	--

Fonte: elaboração própria com base na *National Science Foundation* (2005), Sayão e Sales (2015), Silva (2019) e OpenAIRE (2024).

É fundamental que essas informações estejam incorporadas ao leque de habilidades laborais dos profissionais em Biblioteconomia. Assim como nas tradicionais fontes de informações padronizadas e estruturadas (livros, artigos, dissertações, teses e relatórios), que exigem critérios para o seu tratamento, os dados também possuem suas particularidades e características que devem ser estudadas para a criação de um planejamento para seu adequado tratamento.

5.4 Ciclos de vida dos dados

Para garantir o reuso dos dados além do término da pesquisa, é necessária a realização do processo de curadoria digital para que os dados possam ser coletados, tratados e preservados adequadamente em seu ciclo de vida. Segundo Sales (2014, p. 54), “tanto a gestão quanto o estabelecimento de políticas de preservação fazem parte de um leque de atividades que são englobadas pelo conceito de curadoria digital”, os responsáveis pelo processo da curadoria digital devem elaborar um planejamento que esteja alinhado com as características heterogêneas dos dados. Existe uma grande variação entre dados gerados por diferentes disciplinas, sendo essa uma questão que influenciará fortemente na gestão do seu Ciclo de Vida.

O reconhecimento dessas diferenças torna-se crucial para diversas ações no escopo na gestão de dados de pesquisa e do ciclo de vida da curadoria, posto que cada tipo de dado vai necessitar de processos distintos de gestão e de profundidade de ações de curadoria [...] (Sayão; Sales, 2020, p. 40).

Segundo Higgins (2008), uma curadoria digital praticada levando em consideração os diferentes ciclos de vida dos dados pode promover a existência dos dados a longo prazo, diminuir riscos referentes à obsolescência digital, reduzir esforços na duplicação de outros dados e proporcionar dados com maior qualidade para serem reutilizados.

O processo de curadoria aplicado no ciclo de vida dos dados, de acordo com Plale e Kouper (2017, p. 92), pode ser descrito como: “a vida dos objetos de dados do nascimento à morte”. O processo envolve também a escolha do tempo de permanência que esses dados terão em um sistema eletrônico a curto, médio ou longo prazo. A escolha está estreitamente atrelada à importância que os dados possuem, por exemplo, os dados observacionais que geralmente são preservados por tempo indeterminado por possuírem um valor inestimado para a ciência.

Silva (2019) afirma que o processo de curadoria e preservação digital dos dados vai além da sua inserção em um sistema eletrônico, assim, uma série de medidas deverá ser aplicadas para garantir a autenticidade, integridade, acessibilidade e preservação desses dados para o uso futuro. Neste sentido, para alcançar esses objetivos, é necessária a aplicação de um Modelo de Ciclo de Vida dos Dados de acordo com as necessidades e contexto.

Existem diferentes Modelos de Ciclo de Vida dos Dados (MCVD) que podem ser aplicados para o controle e acompanhamento de cada etapa na gestão dos dados científicos, cada um sendo utilizado de acordo com as necessidades específicas de cada campo do conhecimento científico, embora alguns modelos possuam uma maior abrangência e podem ser utilizados por diferentes disciplinas e campos. Para este propósito, de acordo com Sayão e Sales (2015), Plale e Kouper (2017) e Silva (2019), foram desenvolvidos diversos MCDV, com o propósito de promover a curadoria digital dos dados em diferentes etapas do seu ciclo de vida, em conformidade com as demandas específicas de cada instituição.

A seguir, alguns exemplos serão apresentados, conforme levantamentos realizados por diversos autores (Higgins, 2008; Ball, 2012; Sinaeepourfard *et al.*, 2015; Plale; Kouper, 2017; Silva, 2019):

- *ANDS Data Sharig Service Model* – o MCVD foi desenvolvido pela *Australian National Data Service (ANDS)*, órgão responsável em reunir as informações de provedores de dados e de instituições de investimento público da Austrália. O modelo foi criado especificamente para a curadoria de dados que foram coletados para serem inseridos no sistema dessa instituição. As etapas do Modelo são: Criar, Armazenar, Descrever, Identificar, Registrar, Descobrir, Acessar e Explorar.

- *BLM Model* – o MCVD foi desenvolvido pela *Bureau Of Land Management*, órgão que administra as terras públicas nos EUA. A proposta do Modelo é a realização da curadoria de dados de terras para serem compartilhados entre os usuários do seu sistema. As etapas do Modelo são: Planejar, Adquirir, Manter, Acessar, Avaliar e Arquivar.
- *CSA Model* – o MCVD foi desenvolvido pela *Cloud Security Alliance*, empresa que gerencia ambientes de computação em nuvem. A proposta do Modelo é estabelecer segurança para os dados inseridos em seu sistema de armazenamento em nuvem. As etapas do modelo são: Criar, Armazenar, Usar, Compartilhar, Arquivar e Destruir.
- *DataONE Model* – o MCVD foi desenvolvido pela *Data Network Observation for Earth (DataONE)*. A instituição promove a preservação e o compartilhamento de dados em ciências biológicas e meio ambiente. A proposta do MCVD é estabelecer um padrão de curadoria para os dados disponibilizados pela infraestrutura eletrônica da Instituição. As etapas do Modelo compreendem os processos de Coletar, Assegurar, Descrever, Depositar, Preservar, Descobrir, Integrar e Analisar.
- *Curation Lifecycle Model* – o MCVD foi desenvolvido pela *Digital Curation Centre (DCC)*. A instituição é um centro de especialização líder mundial em curadoria de informações digitais com foco no desenvolvimento de capacidade, competência e habilidades para gerenciamento de dados de pesquisa. A proposta da DCC é estabelecer um modelo que seja mais abrangente na união das ações de curadoria e etapas do ciclo de vida dos dados, seguindo um percurso lógico e flexível. O modelo também proporciona que os criadores de dados e curadores estabeleçam planejamentos mais específicos, pois sua constituição é representada em diferentes camadas para a manipulação dos dados. As camadas estão organizadas da seguinte forma:

Ações do Ciclo de Vida Completo, Ações Sequenciais e Ações Ocasionais. As Ações do Ciclo de Vida Completo são divididas em quatro etapas que são, Descrição e Representação de Informações, Planejamento de Preservação, Vigilância e Participação da Comunidade e Curadoria e Preservação. As Ações Sequenciais fornecem sete etapas que são, Conceitualizar, Criar ou Receber, Avaliar e Selecionar, Ingerir, Ação de Preservação, Armazenar, Acessar, Usar e Reutilizar e Transformar. As Ações Ocasionais incluem Descartar, Reavaliar e Migrar [...] (Ball, 2012, p. 3).

É fundamental que o profissional em Biblioteconomia conheça os diversos modelos existentes para aplicá-los de acordo com o contexto e as características vigentes de cada campo do conhecimento ou tipologia do dado. Segundo Silva (2019), às vezes, nem todas as etapas de um MCVD necessariamente serão utilizadas em um projeto científico, porém, o mais importante é que essa gestão esteja fundamentada em ações que deem significado a esses dados e promova a sua preservação através dos processos de análise. Nessa perspectiva, será alcançado o objetivo maior, que é tornar esses dados intercambiáveis e passíveis de reutilização. São práticas que o profissional bibliotecário deverá incorporar às suas competências laborais com o objetivo de promover uma curadoria digital de forma adequada para permitir o reinício do ciclo de vida dos dados.

5.5 Plano de gestão de dados

O Plano de Gestão de Dados (PGD) é um documento que descreve uma série de informações sobre todos os procedimentos durante a pesquisa científica, relacionados à gestão dos dados e o seu ciclo de vida. Segundo a Fiocruz (2020), o PGD pode ser definido como:

[...] documento formal, dinâmico e atualizável com perguntas orientadoras que estimulam o pesquisador a planejar, de maneira intencional, como se dará a gestão de dados ao longo de todo o seu ciclo de vida - período que é maior que a execução e finalização da pesquisa.

Importante não confundir o Plano de Gestão de Dados com o Plano de Dados Abertos. O primeiro diz respeito ao gerenciamento dos dados especificamente de uma determinada pesquisa, seja ela realizada em qualquer área do conhecimento. Já o segundo se configura em uma política de gerenciamento de dados em uma esfera nacional ou institucional, conforme informações apresentadas na subseção 5.2.

Para organizar as informações contidas em um PGD, diversas instituições criaram modelos de PGD com perguntas prévias que são respondidas pelos responsáveis pela pesquisa. Entre as perguntas que podem compor um modelo de PGD, destacam-se: quem são os responsáveis pelo gerenciamento dos dados; como os dados foram coletados ou criados; que tipos de documentação acompanharão os

conjuntos de dados; qual é o seu volume e tipologias; qual metodologia utilizada para coleta ou criação; quais medidas serão aplicadas para os dados pessoais e sensíveis; onde os dados serão armazenados; quais controles de acesso serão aplicados; quais os requisitos legais, éticos e códigos de conduta e como serão realizados os processos de compartilhamento, curadoria e preservação. (Lima, 2019; Silva, 2019; Príncipe *et al.*, 2020).

Os componentes padrões que formam um modelo de PGD geralmente são organizados em blocos temáticos que são apresentados com uma ou mais perguntas cada, assim, consequentemente, a quantidade de informações presentes em um PGD pode variar de acordo com as perguntas respondidas em decorrência das características de uma área do conhecimento ou do projeto científico.

A necessidade da formulação do PGD pelos pesquisadores está relacionada com uma maior organização na documentação dos processos realizados durante uma pesquisa, mas principalmente com as políticas mandatórias das agências de fomento à pesquisa que condicionam a concessão de financiamento a submissão de um PGD para ser analisado. Em diversas partes do mundo, esta é uma realidade bastante presente, com muitas agências de fomento à pesquisa operando neste sentido, como exposto no Quadro 10.

Quadro 10 - Política mandatória para a formulação do PGD

Agência de Fomento à Pesquisa	País	Política
<i>National Research Foundation (NRF)</i>	África do Sul	Política mandatória para formulação do PGD ⁵⁹
<i>German Research Fundation</i>	Alemanha	Política mandatória para formulação do PGD ⁶⁰
<i>Australian Research Council (ARC)</i>	Austrália	Política mandatória para formulação do PGD ⁶¹
<i>National Health and Medical Research Council (NHMRC)</i>	Austrália	Política mandatória para formulação do PGD ⁶²

⁵⁹ Ver informações em: <https://ufs.libguides.com/c.php?g=977378&p=7067126>

⁶⁰ Ver informações em: <https://www.cms.hu-berlin.de/en/dl-en/dataman-en/work/create-dmp/forschungsdaten-checkliste-en-data.pdf>

⁶¹ Ver informações em: <https://www.arc.gov.au/about-arc/arc-strategies-and-policies/international/research-data-management>

⁶² Ver informações em: <https://www.nhmrc.gov.au/sites/default/files/documents/attachments/Management-of-Data-and-Information-in-Research.pdf>

<i>Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP)</i>	Brasil	Política mandatória para formulação do PGD ⁶³
<i>Canadian Institutes of Health Research (CIHR)</i>	Canadá	Política mandatória para formulação do PGD ⁶⁴
<i>Natural Sciences and Engineering Research Council of Canada</i>	Canadá	Política mandatória para formulação do PGD ⁶⁵
<i>Social Science and Humanities Research Council</i>	Canadá	Política mandatória para formulação do PGD ⁶⁶
<i>National Institute of Health (NIH)</i>	Estados Unidos	Política mandatória para formulação do PGD ⁶⁷
<i>National Science Foundation (NSF)</i>	Estados Unidos	Política mandatória para formulação do PGD ⁶⁸
<i>Netherlands Organization for Scientific Research (NWO)</i>	Holanda	Política mandatória para formulação do PGD ⁶⁹
<i>UKRI</i>	Reino Unido (Inglaterra, Escócia, País de Gales e Irlanda do Norte).	Política mandatória para formulação do PGD ⁷⁰
<i>Welcome Trust</i>	Inglaterra	Política mandatória para a formulação do PGD ⁷¹

Fonte: elaboração própria.

Nas palavras de Spichtinger (2022), o Projeto *Horizon 2020*⁷² foi fundamental para a popularização da obrigatoriedade da formulação do PGD, pois pela primeira vez um órgão com atuação em diversos países estabelecia uma política obrigatoria para esse propósito. O Projeto *Horizon 2020* é uma iniciativa da Comunidade Europeia, com o objetivo de tornar os dados das pesquisas europeias financiadas com

⁶³ Ver informações em: <https://www.abcd.usp.br/apoio-pesquisador/dados-pesquisa/plano-gestao-dados-2/>

⁶⁴ Ver informações em: <https://cihr-irsc.gc.ca/e/54157.html>

⁶⁵ Ver informações em: https://www.nserc-crsng.gc.ca/NSERC-CRSNG/Reports-Rapports/DP/2023-2024/index_eng.asp

⁶⁶ Ver informações em: https://www.sshrc-crsh.gc.ca/funding-financement/apply-demande/guides/guide_preparing_data_management_plan-guide_preparation_plan_gestion_donnees-eng.aspx

⁶⁷ Ver informações em: <https://sharing.nih.gov/data-management-and-sharing-policy/planning-and-budgeting-for-data-management-and-sharing/writing-a-data-management-and-sharing-plan#after>

⁶⁸ Ver informações em: <https://www.nsf.gov/funding/data-management-plan#nsfs-data-sharing-policy-1c8>

⁶⁹ Ver informações em: <https://zenodo.org/records/10143150>

⁷⁰ Ver informações em: <https://library.bath.ac.uk/research-data/data-management-plans/ukri-requirements>

⁷¹ Ver informações em: <https://cms.wellcome.org/sites/default/files/data-management-plans.pdf>

⁷² Ver informações em: https://research-and-innovation.ec.europa.eu/funding/funding-opportunities/funding-programmes-and-open-calls/horizon-2020_en

verba pública em abertos e *FAIR*, com o lema “tão aberto quanto possível, tão fechado quanto necessário”.

As exigências para formulação dos PGD revelam uma crescente preocupação com a preservação dos dados. Nesse sentido, há uma mudança cultural em vias de construção, um contraponto em relação às antigas práticas de descarte dos dados após o final da pesquisa ou a preservação inadequada (armazenamento em suporte físico, *pen drive* ou computador pessoal).

A correta formulação do PGD pode gerar diversos benefícios que envolvem o auxílio na reprodutibilidade da pesquisa, contextualização da criação dos dados permitindo o seu uso posterior, criação de diretrizes para a preservação minimizando a perda de dados, amplia a visibilidade do pesquisador e do estudo e acelera o processo colaborativo entre pesquisadores (Smale *et al.*, 2018; Príncipe *et al.*, 2020).

Um outro benefício gerado é o controle de versionamento dos dados, pois durante a pesquisa diversas versões de dados podem ser geradas, assim os dados que foram concebidos no final da pesquisa podem possuir uma versão totalmente diferente da que foi elaborada no início. O controle de versão permite conferir também transparência no processo da geração dos dados.

O PGD é um documento orgânico, ou seja, é um documento vivo, e, apesar de ser elaborado no início de um projeto de pesquisa, ainda poderá ser atualizado ao longo do processo de desenvolvimento da investigação, ganhando diferentes versões do início ao fim, sempre que necessário. O versionamento, também denominado controle de versões, é uma característica herdada da área de Ciência da Computação e Sistemas de Informação, as quais possuem a prática de nomear e salvar cada nova atualização de softwares e documentos técnicos (Lima, 2019, p. 78).

Com a crescente demanda pela exigência da formulação dos PGD, diversas ferramentas eletrônicas foram construídas contendo modelos de planos para auxiliar os pesquisadores no preenchimento desses documentos, tornando o processo mais rápido e eficiente. As primeiras ferramentas eletrônicas desenvolvidas foram a *DMPonline*⁷³ e a *DMPTool*⁷⁴, respectivamente criadas em 2010 e 2011 (Donnelly; Jones; Pattenden-Fail, 2010; *DMPTool*, 2025). Porém, os pesquisadores atualmente já possuem à sua disposição uma grande variedade de ferramentas com diversas

⁷³ Ver informações em: <https://dmponline.dcc.ac.uk/>

⁷⁴ Ver informações em: <https://dmptool.org/>

características e para diferentes propósitos: *OpenMP*⁷⁵, *easyDMP*⁷⁶, *Argos*⁷⁷, *Data Stewardship Wizard (DSW)*⁷⁸, *Research Data Management Organiser (RDMO)*⁷⁹, *ezDMP*⁸⁰, *DMP Researcher Template for Archaeological Datasets*⁸¹, *DMP Canvas Generator*⁸² e PGD BR⁸³ (Príncipe *et al.*, 2020; Vidal; Gabriel Júnior; Pavão, 2023).

Outra forma de auxiliar os pesquisadores são os serviços de orientação para a construção do PGD desenvolvidos pelas bibliotecas. Os profissionais bibliotecários, por meio de entrevistas realizadas com os pesquisadores, podem identificar lacunas em seus PGD que, futuramente, acarretariam prejuízos para o uso, reproduzibilidade e preservação dos dados (Silva, 2016).

Apesar da relevância do documento para a comunidade científica e a criação de modelos multifacetados, a característica estática dos PGD impossibilita sua interoperabilidade com o ciclo de vida do ecossistema científico. Com o objetivo de transpor essas limitações, uma nova geração de PGD que podem ser acionáveis por máquina está surgindo. O novo mecanismo é intitulado de *machine-actionable Data Management Plans (maDMP)* (Vidal; Gabriel Júnior; Pavão, 2023).

As características interoperáveis dos *maDMP* permitem integrações com o DOI, o ORCID e o ROR, além de estabelecer uma troca de informações de via dupla entre os responsáveis pela criação do *maDMP*, as agências de fomento à pesquisa e os repositórios de dados científicos (Vidal; Gabriel Júnior; Pavão, 2023), tornando o processo de atualização do PGD mais dinâmico e colaborativo. Neste sentido, um estudo desenvolvido no âmbito da *Research Data Alliance (RDA)* elencou dez princípios para que uma ferramenta de PGD possa ser considerada acionável por máquina (Miksa *et al.*, 2019, p. 4).

- Integrar os PGDs com os fluxos de trabalho de todos os intervenientes no ecossistema de dados de investigação;
- Permitir que os sistemas automatizados atuem em representação dos intervenientes no ecossistema de dados de investigação;

⁷⁵ Ver informações em: <https://www.openmp.org/>

⁷⁶ Ver informações em: <https://easydmp.sigma2.no/>

⁷⁷ Ver informações em: <https://argos.openaire.eu/splash/>

⁷⁸ Ver informações em: <https://ds-wizard.org/>

⁷⁹ Ver informações em: <https://rdmorganiser.github.io/en/>

⁸⁰ Ver informações em: <https://ezdmp.org/index>

⁸¹ Ver informações em: <https://vast-lab.org/dmp/ariadneplus/form/>

⁸² Ver informações em: <https://dmp.vital-it.ch/#/login>

⁸³ Ver informações em: <https://pgd.ibict.br/>

- Elaborar políticas que integrem os sistemas e não apenas as pessoas;
- Descrever, tanto para pessoas como para sistemas, os componentes do ecossistema de gestão de dados;
- Utilizar identificadores persistentes e vocabulários controlados;
- Seguir um modelo de dados comum para planos de gestão de dados açãoáveis;
- Disponibilizar PGDs para utilização por parte de pessoas e sistemas;
- Apoiar a avaliação e monitorização da gestão de dados;
- Tornar os PGDs documentos atualizáveis, vivos e com possibilidade de versionamento;
- Tornar os PGDs publicamente disponíveis.

Os PGD podem ser comparados a uma política de gerenciamento dos dados em seu ciclo de vida. Esses documentos contêm as principais informações para possibilitar uma série de procedimentos referentes aos dados em aspectos técnicos, legais e éticos.

5.6 Repositórios de dados científicos

O depósito dos conjuntos de dados em repositórios científicos é um dos procedimentos que podem garantir diversos benefícios aos pesquisadores e à sua pesquisa no que se refere à integridade dos dados, ao compartilhamento, ao aumento de citação e à preservação digital a longo prazo. Essa infraestrutura eletrônica é uma ferramenta estratégica, pois visa garantir um conjunto de serviços em um cenário onde a complexidade inerente no tratamento e na disponibilização dos conjuntos de dados científicos demanda tempo, habilidades e conhecimentos diversificados dos pesquisadores.

Repositórios de dados científicos podem ser definidos como um serviço de informação científica que possibilita o gerenciamento dos dados nas etapas da coleta, do tratamento, da disseminação, da preservação e do reuso dos conjuntos de dados gerados nas pesquisas científicas (Silva, 2016; Sayão; Sales, 2022). Os dados científicos que são depositados em repositórios são constituídos por diferentes tipologias (textos, imagens, vídeos, equações, sequências genéticas), diversificados

formatos (*PDF, Excel, CSV, JSON*) e variados tamanhos (megabytes, gigabytes, petabytes).

Quanto à sua classificação, os repositórios de dados podem ser categorizados como temáticos, institucionais, multidisciplinares ou de uso geral, editoriais e pessoais. Sua classificação é aplicada de acordo com o objetivo a que foi designada, por exemplo: os temáticos abrigam dados de uma determinada disciplina ou área do conhecimento e podem estar associados a uma única instituição ou a várias; os institucionais, como o nome já diz, abrigam apenas dados de uma determinada instituição de ensino ou pesquisa; os multidisciplinares ou de uso geral abrigam dados de diferentes instituições, disciplinas e áreas do conhecimento, podendo ser gerenciados por uma única instituição ou por várias, e os pesquisadores não precisam ser afiliados a instituição mantenedora para realizarem depósitos; os editoriais abrigam dados de periódicos científicos cadastrados na plataforma, enquanto os pessoais abrigam dados de uma única pesquisa e são gerenciados pelos próprios responsáveis pela pesquisa (Silva, 2016; Sayão; Sales, 2022).

Quadro 11 - Classificação dos repositórios de dados científicos

Categorias	Exemplos
Temáticos	<i>GenBank</i> ⁸⁴ , <i>BioGrid</i> ⁸⁵ , <i>Intermagnet</i> ⁸⁶ e <i>IMPC</i> ⁸⁷
Institucionais	<i>Harvard Dataverse</i> ⁸⁸ , <i>Apollo</i> (Repositório da Universidade de Cambridge) ⁸⁹ e <i>Edinburgh DataShare</i> ⁹⁰
Multidisciplinares ou Uso Geral	<i>Deposita Dados</i> ⁹¹ , <i>Zenodo</i> ⁹² e <i>Figshare</i> ⁹³
Editoriais	<i>Dryad</i> ⁹⁴
Próprios	<i>Dropbox</i> ⁹⁵ , <i>Amazon Cloud Drive</i> ⁹⁶ e <i>Microsoft Azure</i> ⁹⁷

Fonte: elaboração própria com base em (Silva, 2016; Sayão; Sales, 2022).

⁸⁴ Ver informações em: <https://www.ncbi.nlm.nih.gov/genbank/>

⁸⁵ Ver informações em: <https://thebiogrid.org/>

⁸⁶ Ver informações em: <https://intermagnet.org/>

⁸⁷ Ver informações em: <https://www.mousephenotype.org/>

⁸⁸ Ver informações em: <https://dataverse.harvard.edu/>

⁸⁹ Ver informações em: <https://www.repository.cam.ac.uk/home>

⁹⁰ Ver informações em: <https://datashare.ed.ac.uk/>

⁹¹ Ver informações em: <https://depositadados.ibict.br/>

⁹² Ver informações: <https://zenodo.org/>

⁹³ Ver informações em: <https://figshare.com/>

⁹⁴ Ver informações em: <https://datadryad.org/>

⁹⁵ Ver informações em: https://www.dropbox.com/pt_BR/

⁹⁶ Ver informações em: <https://aws.amazon.com/pt/products/storage/>

⁹⁷ Ver informações em: <https://azure.microsoft.com/pt-br/>

A instalação de um repositório de dados científicos é um projeto que envolve complexidade e por esse motivo, requer a participação de diferentes setores da instituição. Nesse processo, para garantir a funcionalidade da plataforma, os responsáveis pela sua concepção devem, obrigatoriamente, alinhar as políticas do repositório com a cultura organizacional da instituição, principalmente no âmbito político. Nessa direção, Sayão e Sales (2022, p. 4) discorrem que repositórios de dados se configuram em:

[...] uma proposta de ferramenta em que as atividades, os serviços e os processos que compõem a gestão de dados se agregam e se complementam num ambiente, compreendido como um arcabouço técnico, social e gerencial, em que se efetivam os cuidados com os dados, segundo políticas e diretrizes institucionais definidas para tal.

Obviamente, que a escolha pela instalação de um repositório de dados dependerá também das suas características tecnológicas, existindo diferentes *software* para a escolha, como o *Dspace*, *Dataverse Ckan*, *Fedora* e *EPrints* (Rocha *et al.*, 2021). Por esse motivo, é fundamental a realização de um levantamento de como as diferentes instâncias da instituição se relacionam com os dados científicos, para que estes pressupostos estejam alinhados com as características do *software*. Por exemplo, os serviços prestados por repositórios de dados no contexto das IES invariavelmente enfrentam as adversidades presentes nas diferenças culturais tanto em aspectos disciplinares, quanto nas grandes áreas do conhecimento. Assim, enquanto a física e a área da saúde possuem uma relação mais proeminente no que diz respeito ao compartilhamento e tratamento de dados, outras disciplinas e áreas ainda carecem de um desenvolvimento neste aspecto (Pinfield; Cox, 2014).

As diferenças disciplinares e das grandes áreas do conhecimento se manifestam também nas diferentes tipologias de dados que são geradas. Essa é uma questão sensível que exige da gestão do repositório uma equipe multidisciplinar, metadados que descrevam adequadamente diferentes tipologias de dados e um modelo de ciclo de vida de dados abrangente. Outro aspecto importante diz respeito aos dados pessoais e sensíveis. Nesse sentido, é imprescindível que o ciclo de vida dos dados e a política adotada pela gestão do repositório estejam em consonância com as leis de proteção de dados do seu respectivo país. Essas práticas visam garantir credibilidade ao repositório e a proteção de informações que não podem ser disponibilizadas para o acesso público.

Em 2018, um novo regulamento sobre dados pessoais e sensíveis, que abrange toda a União Europeia, entrou em vigor. O regulamento, intitulado *General Data Protection Regulation (GDPR)*, foi previsto para servir de referência na legislação do processamento e compartilhamento de dados pessoais e sensíveis nas pesquisas científicas, históricas, estatísticas e de saúde (OpenAIRE, 2018).

No Brasil, a Lei Geral de Proteção de Dados (LGPD) é o principal documento que determina quais procedimentos serão adotados ao tratamento e compartilhamento de dados pessoais e sensíveis de pessoa natural ou pessoa jurídica de direito público ou privado, com determinações que se estendem para os dados em suporte físico e digital (Brasil, 2018).

Os dados pessoais e sensíveis armazenados em um repositório e que passaram pelos procedimentos para evitar a identificação de uma pessoa, ou seja, técnicas de anonimização, não estão sujeitos a aplicação da LGPD. Contudo, é fundamental que o responsável pela anonimização dos dados pessoais e sensíveis tenha pleno conhecimento da técnica, pois a sua aplicação incorreta acarretará a transgressão da LGPD.

Ressalta-se que o dado somente é considerado anonimizado se não permitir que, por meios técnicos ou outros, seja reconstruído o caminho para revelar quem é o(a) titular do dado. Se a identificação ocorrer, não se tratará de dado anonimizado, mas sim de dado pseudonimizado, e estará sujeito à LGPD (Ministério do Esporte, 2025).

A tabela a seguir demonstra as aplicações da LGPD para o tratamento e compartilhamento dos dados pessoais e sensíveis que pode servir como referência para os repositórios no ciclo de vida dos dados.

Quadro 12 – Tratamento dos dados pessoais e sensíveis pela LGPD

CICLO DE VIDA DOS DADOS	
Fase do ciclo	LGPD
Coleta	Os dados coletados devem obedecer ao princípio da necessidade e da finalidade.
Processamento	O processamento de dados só poderá ser realizado se o tratamento estiver enquadrado no Art. 7º da LGPD.
Análise	A análise dos dados deve levar em consideração a finalidade da coleta. Devem ser obedecidos os

	princípios de tratamento com propósito legítimo específico e explícito.
Compartilhamento	O compartilhamento de dados deve ser consentido pelos seus titulares* *Ver Inciso II do Art. 3º do Decreto nº 10.046/2019 ⁹⁸ .
Armazenamento	Os dados pessoais devem ser armazenados e mantidos por prazos definidos, ou seja, até que a finalidade seja alcançada ou deixem de ser necessários ou pertinentes ao alcance da finalidade.
Reutilização	Um novo consentimento deve ser solicitado sempre que houver mudança de finalidade.
Eliminação	Os dados pessoais devem ser eliminados após o término de seu tratamento.

Fonte: Barbosa; Lyra (2019).

Tratando-se ainda de aspectos legais, porém no reuso dos dados, a aplicação de licenças de uso em repositórios se configura em uma boa prática diante das múltiplas legislações relacionadas aos direitos autorais em diversos países. No Brasil, a Lei nº 9.610 (Brasil, 1998) é o principal documento jurídico que contém determinações sobre o tema.

As licenças de uso permitem que os usuários identifiquem se os dados podem ser reutilizados e para quais propósitos podem ser utilizados. Nesse cenário, segundo Ball (2014, p. 3), licenças de uso de dados aplicadas em repositórios são mecanismos que “permite ao detentor de direitos autorizar uma segunda parte a realizar ações que, de outra forma, infringiriam os direitos detidos”. Para esse fim, existem diversos tipos de licenças, porém as mais utilizadas são as *Creative Commons* e *Open Data Commons*.

As licenças *Creative Commons* eram utilizadas frequentemente para licenciar livros, manuais, guias, *slides*, vídeos, fotos e políticas, mas com a ascensão do compartilhamento de dados passaram a ser utilizadas para licenciar também os dados. As licenças da *Creative Commons* são formadas pelos princípios da Atribuição, Sem Derivações, Não Comercial e Compartilha Igual. Quando associadas, permitem a criação de seis licenças, existindo ainda a Marca de Domínio Público e o Creative Commons Zero (*Creative Commons* Brasil, 2020; *Creative Commons*, 2025):

⁹⁸ Ver informações em: https://www.planalto.gov.br/ccivil_03/_ato2019-2022/2019/decreto/d10046.htm

- *CC BY*: Esta licença permite que reutilizadores distribuam, remixem, adaptem e criem a partir do material em qualquer meio ou formato, desde que seja dada a devida atribuição ao criador. A licença permite o uso comercial;
- *CC BY-SA*: Esta licença permite que reutilizadores distribuam, remixem, adaptem e criem a partir do material em qualquer mídia ou formato, desde que seja dada a devida atribuição ao criador. A licença permite o uso comercial. Se você remixar, adaptar ou criar a partir do material, deverá licenciar o material modificado sob termos idênticos;
- *CC BY-NC*: Esta licença permite que reutilizadores distribuam, remixem, adaptem e criem a partir do material em qualquer meio ou formato, apenas para fins não comerciais, desde que seja dada a devida atribuição ao criador;
- *CC BY-NC-SA*: Esta licença permite que reutilizadores distribuam, remixem, adaptem e criem a partir do material em qualquer meio ou formato, apenas para fins não comerciais, desde que seja dada a devida atribuição ao criador. Se você remixar, adaptar ou criar a partir do material, deverá licenciar o material modificado sob os mesmos termos;
- *CC BY-ND*: Esta licença permite que reutilizadores copiem e distribuam o material em qualquer meio ou formato, apenas em formato não adaptado, desde que seja dada a devida atribuição ao criador. A licença permite o uso comercial;
- *CC BY-NC-ND*: Esta licença permite que reutilizadores copiem e distribuam o material em qualquer meio ou formato, apenas em formato não adaptado, apenas para fins não comerciais e desde que seja dada a devida atribuição ao criador;
- Marca de Domínio Público: é um sinal distintivo que permite identificar facilmente que o período de proteção de uma obra já transcorreu, e que, portanto, ela já se encontra em domínio público. Ela é muito utilizada por instituições de memória como museus, bibliotecas e arquivos para divulgar obras de sua coleção e comunicar ao público que elas não são mais protegidas pelo direito autoral e podem ser livremente usadas;
- *CC O*: o zero é um instrumento jurídico por meio do qual uma pessoa dedica sua obra ao domínio público, renunciando a todos os direitos autorais sobre a obra, na medida do permitido por lei. Esse instrumento permite que se copie,

modifique, distribua e execute a obra, inclusive para fins comerciais. No Brasil, como não é possível renunciar aos direitos morais de autor, é preciso sempre indicar a autoria da obra, mesmo que esteja em CC0. Isso não invalida a licença nem seu uso, inclusive porque o licenciamento se dá de forma global, e outros países podem permitir renúncias a direitos que o Brasil não permite.

Segundo Ball (2014), embora os princípios das Licenças CC permaneçam inalterados ao longo dos anos, os seus textos sofreram algumas alterações para se adaptar a diferentes contextos legais, originando assim diferentes versões. Diante o exposto, o autor afirma que a versão 4 é a mais indicada para o uso em dados e banco de dados, pois possui indicações mais específicas para adaptar-se a diferentes legislações.

Já o *Open Data Commons* foi criado pela *Open Knowledge Foundation (OKF)*, com o objetivo específico de licenciar dados e conjuntos de dados em base de dados. O *Open Data Commons* está dividido em três tipos de licenças:

- *Attribution License (ODC-BY)*: é um contrato de licença que visa permitir que os usuários compartilhem, modifiquem e usem livremente um banco de dados, conforme especificações da *ODC-BY*⁹⁹;
- *Open Database License (ODbL)*: é um contrato de licença que visa permitir que os usuários compartilhem, modifiquem e utilizem livremente um banco de dados, conforme as especificações da *ODbL*¹⁰⁰;
- *Public Domain Dedication and License (PDDL*¹⁰¹): coloca os dados ou a base de dados em domínio público renunciando a todos os direitos (OKF, 2024c).

Considerando ainda as boas práticas em repositório de dados científicos, os responsáveis pela instalação da plataforma devem conhecer instrumentos que possibilitem que os dados possam ser localizáveis, acessíveis, interoperáveis e reusáveis, tanto por humanos quanto por máquinas. Nesse sentido, a aplicação dos Princípios FAIR pode representar uma importante ferramenta para alcançar esses parâmetros.

⁹⁹ Ver informações em: <https://opendatacommons.org/licenses/by/1-0/>

¹⁰⁰ Ver informações em: <https://opendatacommons.org/licenses/odbl/summary/>

¹⁰¹ Ver informações em: <https://opendatacommons.org/licenses/pddl/1-0/>

Na perspectiva da confiabilidade do repositório, os gestores podem adequar-se aos critérios de avaliações como *Core Trust Seal*¹⁰², *ISO 16363*¹⁰³ e *Nestor*¹⁰⁴, com o objetivo de obter selos de certificações de repositórios de dados confiáveis (Wilksom *et al.*, 2016; Rocha *et al.*, 2021). Os selos de certificação são concedidos aos repositórios que alcançaram uma série de requisitos em sua avaliação. Os critérios da *Core Trust Seal*, por exemplo, envolvem a avaliação da infraestrutura organizacional, gerenciamento de objetos digitais, tecnologia da informação e segurança (*Core Trust Seal*, 2022).

No aspecto da confiabilidade, está prevista também a preservação digital, que confere o acesso, compartilhamento, recuperação e preservação dos objetos digitais a longo prazo em um repositório. De acordo com Márdero-Arellano (2012), o modelo de Preservação Digital Distribuída é o mais indicado para promover a segurança dos objetos digitais, pois consiste em manter uma cópia do objeto digital armazenado geograficamente em locais diferentes, assegurando assim que os objetos digitais poderão ser recuperados caso ocorra algum ataque cibernético ou desastre natural.

Diante deste complexo cenário, a criação de serviços e produtos para orientar os pesquisadores na escolha de repositórios para o depósito dos seus conjuntos de dados pode ser bastante elucidativa. Os bibliotecários, com os seus conhecimentos de pesquisa estruturada, podem participar desse processo de forma ativa na seleção de repositórios que possam proporcionar maior visibilidade aos dados, levantamentos dos pré-requisitos para aceitação dos dados nesses repositórios e a criação de guias, manuais, tutoriais e vídeos para orientar os pesquisadores (Silva, 2016).

O *re3data* é uma ferramenta que pode ser bastante útil nesse processo de descoberta de repositórios de dados científicos. A plataforma é um diretório que possui 3.360 repositórios de dados de diversos países cadastrados em sua base, com informações sobre as suas categorias, instituições mantenedoras e tipo de *software* que utilizam. Os dez países com o maior quantitativo de repositórios cadastrado no Diretório são Estados Unidos (1200), Alemanha (542), Canadá (403), Reino Unido (325), França (139), Australia (107), China (86), Suíça (86), Holanda (83) e Espanha (63), respectivamente (*re3data*, 2025).

¹⁰² Ver informações em: <https://www.coretrustseal.org/why-certification/requirements/>

¹⁰³ Ver informações em: <http://www.iso16363.org/>

¹⁰⁴ Ver informações em: <https://www.dcc.ac.uk/resources/repository-audit-and-assessment/nestor>

Conforme a pesquisa realizada por Pavão *et al.* (2018), que utilizou três fontes de informação, incluindo o *re3data*, o quantitativo de repositórios de dados científicos no Brasil em 2018 era de quinze, sendo quatro de parcerias com instituições internacionais. Já em março de 2025, através de dados extraídos do *re3data* pelo autor desta presente pesquisa, o Brasil alcançou o quantitativo de vinte três repositórios de dados científicos, sendo sete proveniente de parcerias estabelecidas com outros países (*re3data*, 2025). Os dados demonstram que de 2018 a 2025 apenas oito repositórios foram instalados no país. Assim, neste cenário, o quantitativo de instalação foi de praticamente um por ano. Os dados ainda revelam uma discrepância na quantidade de repositórios entre o Brasil e os dez primeiros no *re3data*. Importante ressaltar que neste quantitativo não estão presentes as quatro instituições que foram contempladas no Edital para incubação de Repositórios de Dados de Pesquisa¹⁰⁵, lançado em 2021.

O referido Edital nasce do acordo tripartite de cooperação entre RNP, IBICT e CNPq, mais o suporte consultivo do Grupo de Trabalho Rede de Dados de Pesquisa Brasileira (GT-RDP Brasil). Essa iniciativa visava acelerar o processo de instalação de repositórios de dados científicos no Brasil, com a seleção de quatro instituições para transferência de conhecimento em gestão de repositórios, incubação dos repositórios pelo período de 9 meses e fornecimento de identificadores persistentes (DOI) limitado a 10.000 registros anuais. As instituições selecionadas foram Universidade Federal do Ceará (UFC), Universidade Federal Rural da Amazônia (UFRA), Centro Brasileiro de Pesquisa Físicas (CBPF) e Universidade Federal de Goiás (UFG). Nesse seguimento, a execução do Edital visava cumprir o 4º Plano de Ação Nacional em Governo Aberto em seu Compromisso 3 no Marco 7 (Ciuffo; Felicíssimo, 2025).

Ainda falando sobre o Edital, a seleção das instituições foi realizada através da aplicação do documento **Conjuntos de Indicadores para Aferição do Grau de Maturidade de Abertura dos Dados Científicos**, criado no 4º Plano de Ação Nacional em Governo Aberto em seu Compromisso 3 no Marco 9. O documento está estruturado em quatro eixos: Governança; Gestão de Dados Científicos; Cultura Organizacional e Infraestrutura Tecnológica. Cada eixo possui um conjunto de

¹⁰⁵ Ver informações em: <https://www.gov.br/ibict/pt-br/central-de-conteudos/noticias/2021/julho2021/ibict-cnpq-e-rnp-lancam-edital-para-a-criacao-de-repositorios-de-dados-de-pesquisa>

perguntas com o objetivo de identificar o grau de maturidade das práticas em abertura de dados em uma instituição (Parceria para o Governo Aberto, 2010).

Oportunamente, é importante citar outra iniciativa para acelerar a instalação de repositórios de dados científicos no Brasil, denominada Rede Brasileira de Re却t{titórios Digitais (RBRD). A Rede foi fundada pelo IBICT em 2014 com a nomenclatura de Rede Brasileira de Re却t{titórios Institucionais de Publicações Científicas em Acesso Aberto (RIAA). Posteriormente, para incorporar ações em prol do desenvolvimento de repositórios de dados científicos, o seu nome foi modificado para RBRD (Almeida; Sánchez-Tarragó, 2023).

Uma rede de repositórios pode gerar conexões entre as comunidades, aumentando assim a interface entre fontes de dados de diferentes disciplinas e em repositórios específicos ou multidisciplinares (Pavão; Rocha; Gabriel Júnior, 2018, p. 331).

A RBRD é formada por cinco sub-redes: Rede Norte, Rede Sul, Rede Sudeste, Rede Nordeste e Rede Centro-Oeste, com coordenações e regimentos próprios. As sub-redes desempenham um importante papel na disseminação de informações estratégicas para o fortalecimento da Ciência Aberta no Brasil, com a organização de eventos, palestras, cursos, treinamentos, grupos de trabalho para compartilhar boas práticas e tirar dúvidas na instalação e gerenciamento de repositórios institucionais de publicações científicas, portais institucionais de periódicos científicos e repositórios institucionais de dados científicos (Almeida; Sánchez-Tarragó, 2023). Na estrutura da RBRD, o Núcleo de Dados de Pesquisa (NDP) é o principal grupo responsável em criar mecanismos para abertura de dados e apoio na instalação de repositórios de dados científicos nas instituições membros (RNP, 2024).

A RBRD se configura em uma iniciativa estratégica para abertura de dados e crescimento de repositórios de dados científicos nas IES do Brasil, pois funciona de forma colaborativa com o compartilhamento de informações e boas práticas dos seus respectivos membros, ao mesmo tempo que desenvolve ações de forma descentralizada, nessa perspectiva promovendo celeridade em suas ações e o amadurecimento das práticas da Ciência Aberta no país.

6 PERCURSO METODOLÓGICO

A pesquisa científica é o sustentáculo da existência da ciência e a sua aplicação permite descobrir e interpretar os fenômenos existentes entre os objetos de estudo, ao mesmo tempo que evidencia novas possibilidades de investigação desses mesmos objetos. Nesse percurso, cria-se a possibilidade do surgimento de novas pesquisas para o funcionamento cíclico da ciência. Para tanto, faz-se necessário selecionar métodos de investigação para condução desse propósito.

Nesse sentido, delineou-se esta seção, com o objetivo de percorrer o seguinte percurso: **caracterização da pesquisa; escolha dos procedimentos técnicos; definição do universo; critérios empregados para a seleção da amostra; instrumentos de coleta de dados; coleta de dados e a análise dos resultados**. A passagem por este percurso será obrigatoriamente necessária para alcançar o **objetivo geral**: Diagnosticar como os sistemas de bibliotecas das universidades federais do Nordeste brasileiro estão promovendo ou se preparando para promover serviços e produtos para gestão dos dados científicos das suas instituições. Para alcançar o objetivo geral, foram eleitos os seguintes **objetivos específicos** (OE):

- OE1: mapear a partir da literatura científica, quais os serviços e produtos para o gerenciamento de dados científicos;
- OE2: averiguar nos *sites* de bibliotecas de instituições de ensino superior estrangeiras a existência de serviços e produtos destinados ao gerenciamento de dados científicos, contribuindo com informações complementares para elaboração do diagnóstico;
- OE3: identificar em qual estágio se encontram os produtos e serviços para gestão de dados científicos nos sistemas de bibliotecas das universidades do Nordeste brasileiro.

Dito isto, a presente pesquisa caracteriza-se, **quanto ao objetivo**, como **descritiva**, pois o presente estudo propõe descrever os fatos e fenômenos do objeto investigado, com a intenção de criar elementos para uma melhor compreensão do tema abordado. A pesquisa descritiva também tem como objetivo investigar as características que compõem um determinado grupo social, porém não possui a

intenção de alterar ou intervir nos aspectos do objeto estudado e sim descobrir e observar este fenômeno para interpretá-lo (Gil, 2010; Rudio, 2007).

A **abordagem** da pesquisa é classificada como **qualitativa e quantitativa**, pois, da perspectiva qualitativa, orienta-se com o objetivo de estudar os elementos que constituem as relações sociais do grupo e de documentos em sua atuação e no seu contexto de um ponto de vista holístico, com a intenção de investigar, analisar e interpretar fatos, para assim proporcionar uma variedade de modelos para a investigação do objeto (Córdova, 2009; Severino, 2013). No que se refere ao ponto de vista quantitativo, a pesquisa propõe apresentar análises a partir da quantificação das práticas e planejamentos da amostra em estudo. A aplicação da abordagem qual-quant proporciona diversas vantagens para o pesquisador, podendo suprir lacunas na abordagem quantitativa para alcançar resultados robustos.

Quanto aos **procedimentos técnicos**, a presente investigação é classificada como revisão de literatura, documental e de levantamento. Marconi e Lakatos (2003, p. 221) esclarecem que os procedimentos técnicos “constituem etapas mais concretas da investigação, com finalidade mais restrita em termos de explicação geral dos fenômenos menos abstratos. Pressupõem uma atitude concreta em relação ao fenômeno e estão limitadas a um domínio particular”.

6.1 Universo e amostra da pesquisa

A presente investigação tem como universo as bibliotecas das universidades federais do Nordeste brasileiro. De acordo com Lakatos e Marconi (2003, p. 223), “[...] universo ou população é o conjunto de seres animados ou inanimados que apresentam pelo menos uma característica em comum”. A escolha pelas universidades federais é fundamentada por estas instituições produzirem a maioria das publicações científicas no país (Moura, 2019) e, consequentemente, a maioria dos dados científicos.

Nesse sentido, pressupõe-se como imprescindível que as referidas instituições realizem planejamento para gerenciar os dados científicos dessas produções com a intenção de promover o seu acesso, compartilhamento e preservação à longo prazo. O gerenciamento de dados científicos contribui consideravelmente para o desenvolvimento da ciência, promove a visibilidade dos pesquisadores, das suas

pesquisas e das instituições. Além disso, o gerenciamento de dados científicos converge com a proposta da Ciência Aberta, a qual orienta que as produções científicas realizadas com a destinação de verba pública precisam ser compartilhadas de forma irrestrita para a comunidade científica e a sociedade em geral, salvo em casos específicos como é o caso dos dados pessoais e sensíveis, em que são aplicadas algumas restrições.

Quanto à escolha pela Região Nordeste do Brasil, ocorre pela participação do autor da presente pesquisa como membro da Rede Nordeste de Repositórios Digitais (Renere) e, consequentemente, pelo seu interesse no desenvolvimento da gestão de dados científicos nesta região.

Tendo em vista que o universo era muito amplo, decidiu-se por investigar uma amostra de bibliotecas das universidades federais do Nordeste brasileiro. A seleção da amostra foi dividida em duas etapas, sendo a primeira etapa composta por todas as universidades federais dos nove estados da Região Nordeste: Alagoas, Bahia, Ceará, Maranhão, Paraíba, Pernambuco, Piauí, Rio Grande do Norte e Sergipe.

Para obter os dados desta amostra, foi utilizado o método de observação sistemática, que, conforme Lakatos e Marconi (2017), é utilizado para levantamentos de dados em um cenário que necessite de um certo controle para responder aos propósitos preestabelecidos, para assim diminuir a influência do pesquisador sobre o processo.

O levantamento foi realizado no *site* do e-MEC¹⁰⁶, sistema em que as Instituições de Ensino Superior (IES) do Brasil são cadastradas com a intenção de criar um banco de dados com diversas informações sobre as referidas instituições. No sistema, utilizaram-se a busca avançada e aplicaram-se quatro filtros para o refinamento da busca:

- a) no filtro **Busca**, a opção “Instituição de Ensino Superior” foi a selecionada;
- b) no filtro **Unidade Federativa (UF)**, foi inserido o Estado para obter a informação da quantidade de IES que esta UF possuía;
- c) no filtro **Categoria Administrativa**, a opção “Pública Federal” foi a escolhida;
- d) no filtro **Organização Acadêmica**, a opção “Universidade” foi a selecionada.

Ao final do levantamento, recuperou-se o quantitativo de 20 universidades, dispostas no quadro a seguir:

¹⁰⁶ Ver informações em: <https://emec.mec.gov.br/emec/nova>

Quadro 13 – Primeira fase da seleção da amostra da pesquisa

Estado	Sigla	Nome
Alagoas	UFAL	Universidade Federal de Alagoas
Bahia	UFBA	Universidade Federal da Bahia
Bahia	UFOB	Universidade Federal do Oeste da Bahia
Bahia	UFRB	Universidade Federal do Recôncavo da Bahia
Bahia	UFSB	Universidade Federal do Sul da Bahia
Ceará	UNILAB	Universidade da Integração Internacional da Lusofonia Afro-Brasileira
Ceará	UFCA	Universidade Federal do Cariri
Ceará	UFC	Universidade Federal do Ceará
Maranhão	UFMA	Universidade Federal do Maranhão
Paraíba	UFPB	Universidade Federal da Paraíba
Paraíba	UFCG	Universidade Federal de Campina Grande
Pernambuco	UNIVASF	Universidade Federal do Vale do São Francisco
Pernambuco	UFPE	Universidade Federal de Pernambuco
Pernambuco	UFAPE	Universidade Federal do Agreste de Pernambuco
Pernambuco	UFRPE	Universidade Federal Rural de Pernambuco
Piauí	UFDPAR	Universidade Federal do Delta do Parnaíba
Piauí	UFPI	Universidade Federal do Piauí
Rio Grande do Norte	UFRN	Universidade Federal do Rio Grande do Norte
Rio Grande do Norte	UFERSA	Universidade Federal Rural do Semi-Árido
Sergipe	UFS	Universidade Federal de Sergipe

Fonte: elaboração própria com base no Cadastro e-MEC (2024).

Na segunda etapa, o critério de seleção foram as universidades que possuíam, no período da coleta de dados, sistema de bibliotecas e repositório institucional de publicações científicas. A existência dessas duas estruturas como critério para fazer parte da amostra justifica-se porque:

- entende-se que as universidades que possuem sistema de bibliotecas gozam de uma estrutura administrativa com maiores capacidades de oferecer serviços de forma sistematizada e a longo prazo;
- a presença de um repositório institucional de publicação científica instalado na instituição serve como ponto de partida para adentrar no universo do gerenciamento de dados científicos. Um exemplo dessa afirmação foi o edital “Incubação de Repositórios de Dados de Pesquisa” lançado pelo IBICT, RNP e CNPq em 2021, no qual constava como

critério de desempate para a seleção das instituições que participaram da proposta a existência de repositório institucional de publicação científica instalado na universidade.

Esta etapa da seleção da amostra foi realizada no buscador do Google, com todas as instituições selecionadas na primeira etapa. Para realizar a consulta no Google, foi empregada como estratégia de busca a palavra-chave “Sistema de Bibliotecas” juntamente com o sinal “+” e a sigla da universidade (ex.: Sistema de Bibliotecas+UFAL). O objetivo da busca era identificar nos *sites* das bibliotecas quais universidades possuíam sistema de bibliotecas e repositório institucional de publicação científica.

Nos *sites* das instituições que não continham informações sobre a existência de repositório, realizou-se uma segunda pesquisa no buscador do Google, empregando como estratégia de busca a palavra-chave “Repositório Institucional” juntamente com o sinal “+” e a sigla da universidade (ex.: Repositório Institucional+UFAL). Ao término da consulta no buscador do Google, a amostra da pesquisa ficou composta por quatorze universidades que atenderam aos dois critérios da seleção da amostra (Quadro 14).

Quadro 14 – Segunda fase da seleção da amostra da pesquisa

ESTADO	SIGLA	NOME	SISTEMA DE BIBLIOTECAS	REPOSITÓRIO
Alagoas	UFAL	Universidade Federal de Alagoas	Sim	Sim
Bahia	UFBA	Universidade Federal da Bahia	Sim	Sim
Bahia	UFRB	Universidade Federal do Recôncavo da Bahia	Sim	Sim
Ceará	UNILAB	Universidade da Integração Internacional da Lusofonia Afro-Brasileira	Sim	Sim
Ceará	UFC	Universidade Federal do Ceará	Sim	Sim

Maranhão	UFMA	Universidade Federal do Maranhão	Sim	Sim
Paraíba	UFPB	Universidade Federal da Paraíba	Sim	Sim
Paraíba	UFCG	Universidade Federal de Campina Grande	Sim	Sim
Pernambuco	UFPE	Universidade Federal de Pernambuco	Sim	Sim
Pernambuco	UFRPE	Universidade Federal Rural de Pernambuco	Sim	Sim
Piauí	UFPI	Universidade Federal do Piauí	Sim	Sim
Rio Grande do Norte	UFRN	Universidade Federal do Rio Grande do Norte	Sim	Sim
Rio Grande do Norte	UFERSA	Universidade Federal Rural do Semi-Árido	Sim	Sim
Sergipe	UFS	Universidade Federal de Sergipe	Sim	Sim

Fonte: elaboração própria.

6.2 Instrumentos e procedimentos da coleta de dados

Para alcançar o OE1, foi realizado um mapeamento das produções científicas com o objetivo de identificar oferta de serviços e produtos para gestão de dados científicos (SPGDC) no Apêndice D.

No que se refere ao OE2, utilizou-se o procedimento Revisão Sistemática para identificar SPGDC em *sites* de bibliotecas estrangeiras. Para proceder com a seleção, recorreu-se a três *rankings* internacionais de universidades em sua edição 2024, a saber: *Academic Ranking of World Universities (ARWU)*¹⁰⁷, *QS Top Universities*¹⁰⁸ e

¹⁰⁷ Ver informações em: <https://www.shanghairanking.com/rankings/arwu/2024>

¹⁰⁸ Ver informações em: https://www.topuniversities.com/world-university-rankings?tab=indicators&sort_by=rank&order_by=asc

o *Times Higher Education (THE)*¹⁰⁹, para selecionar a amostra das universidades estrangeiras. Definiram-se como critérios de seleção para composição dessa amostra:

- a universidade deveria aparecer simultaneamente nos três *rankings* até a quinta colocação;
- a universidade selecionada deveria ofertar SPGDC por suas bibliotecas.

Esclarece-se que, para realizar a revisão sistemática, utilizou-se um protocolo disponível no Apêndice C.

Quadro 15 – Universidades estrangeiras selecionadas em ranking internacionais

<i>Academic Ranking of World Universities (ARWU)/Shanghai Ranking</i>	<i>QS Top Universities</i>	<i>Times Higher Education (THE)</i>
1 – Harvard University	1 – Massachusetts Institute of Technology (MIT)	1 – University of Oxford
2 – Stanford University	2 – Imperial College London	2 – Stanford University
3 – Massachusetts Institute of Technology (MIT)	3 – University of Oxford	3 – Massachusetts Institute of Technology (MIT)
4 – University of Cambridge	4 – Harvard University	4 – Harvard University
5 – University of California, Berkeley	5 – University of Cambridge	5 – University Of Cambridge

Fonte: elaboração própria com base no *Academic Ranking of World Universities (ARWU)/Shanghai Ranking*, *QS Top Universities* e *Times Higher Education (THE)*.

Esses *rankings* definem, para classificação das universidades, critérios, a exemplo¹¹⁰ de: pesquisadores altamente citados, artigos publicados na *Nature* e na *Science*, artigos indexados na *Web of Science*, ensino, pesquisa e descoberta, experiência de aprendizagem e ambiente de pesquisa, patentes e qualidade da pesquisa. Ao final, foram selecionadas: *Harvard University*, *Massachusetts Institute of Technology (MIT)* e a *University of Cambridge*.

¹⁰⁹ Ver informações em: <https://www.timeshighereducation.com/world-university-rankings/2024/world-ranking>

¹¹⁰ Ver informações em:

<https://www.shanghairanking.com/methodology/arwu/2024>

<https://www.topuniversities.com/world-university-rankings/methodology>

<https://www.timeshighereducation.com/world-university-rankings/world-university-rankings-2025-methodology>

Sobre o OE3, foi realizada a aplicação de um questionário para as coordenações e diretorias responsáveis pelos sistemas de bibliotecas das quatorze instituições selecionadas para fazer parte da amostra. A aplicação do questionário objetivou diagnosticar como os sistemas de bibliotecas das universidades federais do Nordeste brasileiro estão promovendo ou se preparando para promover SPGDC em suas respectivas instituições.

Para a construção do referido instrumento de coleta de dados, consultaram-se três fontes de informações, tais como:

- a) literatura científica, a partir da perspectiva de autores que discutem gestão de dados científicos;
- b) serviços e produtos investigados em *sites* de bibliotecas de universidades estrangeiras;
- c) perguntas presentes no documento “Conjunto de Indicadores para Aferição do Grau de Maturidade de Abertura dos Dados Científicos”, as quais foram adaptadas para o questionário desta pesquisa¹¹¹.

Ao final do processo foi formulado um questionário contendo doze perguntas fechadas e uma aberta, o qual foi enviado para o e-mail das coordenações e diretorias dos sistemas de bibliotecas. O instrumento de coleta de dados esteve aberto para resposta no período de 18 de fevereiro até 03 de abril de 2025.

A taxa de resposta do questionário foi de 100%, proporcionando enriquecimento e validade para análise dos resultados. A participação do autor da pesquisa como membro da Rede Nordeste de Repositórios Digitais possibilitou o acesso a diversos membros dos sistemas de bibliotecas dessa Região, configurando-se como principal motivo para o alcance da referida taxa de resposta.

6.3 Pré-teste

Durante a realização de uma pesquisa, a etapa de pré-teste é fundamental para solução de possíveis problemas que venham a ser identificados no instrumento de coleta de dados. Segundo Gerhardt e Silveira (2009), faz-se necessária a realização do pré-teste do questionário com uma pequena população com o objetivo de identificar possíveis inconsistências na sua formulação. Corroborando com as autoras, a

¹¹¹ Informações mais detalhadas sobre este documento foram apresentadas na subseção 5.6.

validação do questionário foi realizada com dois sistemas de bibliotecas de universidades federais do Nordeste brasileiro que foram descartadas por não atenderem aos critérios da seleção da amostra, na segunda etapa. O pré-teste proporcionou a identificação de lacunas na formulação de algumas perguntas do questionário, pois possuía a seguinte estrutura para todas as perguntas, conforme a Figura 5.

Figura 5 – Pergunta da primeira versão do questionário

2 - O sistema de bibliotecas possui Política/Norma implementada para gestão dos dados científicos dos pesquisadores da universidade?

a) Possui.

b) Não possui, mas prevê planejamento para sua formulação.

c) Não possui e nem prevê planejamento para sua formulação.

Fonte: elaboração própria (2025).

Como foi estruturada, a pergunta permitia ao respondente optar pela resposta “a”, mesmo não possuindo política/norma, serviço ou produto para a GDC. Esse mesmo problema foi detectado em outras perguntas, exceto na pergunta 13, que era aberta. Para solucionar a questão, foi adicionada uma resposta condicional para que o respondente, ao marcar a opção “a”, complementasse a resposta com o envio do documento através do *link* disponível no próprio questionário ou para o e-mail do pesquisador, representado pela Figura 6.

Figura 6 – Pergunta da versão final do questionário

2 - O sistema de bibliotecas possui Política/Norma implementada para gestão dos dados científicos dos pesquisadores da universidade? *

a) Possui.

b) Não possui, mas prevê planejamento para sua formulação.

c) Não possui e nem prevê planejamento para sua formulação.

2.1 - Caso tenha respondido "possui" na pergunta anterior, por favor inserir o link da Política/Norma a baixo ou envie o pdf para o email uillisassis@hotmail.com.

Texto de resposta longa

Fonte: elaboração própria (2025).

Ao término dos ajustes, o questionário obteve uma estrutura diferente do pré-teste para as perguntas compreendidas entre o número 2 (dois) ao 12 (doze). Esclarece-se que as respostas condicionais referentes às respostas “a”, cujo respondente não comprovou a existência do serviço ou produto com o envio do documento, foram consideradas como uma resposta “c”.

Destaca-se que, em caso das respostas referentes à pergunta 11.1 não fossem elucidativas, realizavam-se verificações no *site* da instituição, para a constatação do referido serviço. Optou-se por realizar a verificação somente nesta pergunta porque, durante a realização do pré-teste, alguns respondentes confundiram “repositório institucional de publicação científica” com “repositório Institucional de dados científicos”.

6.4 Análise dos dados

A análise dos dados referente ao OE2 ocorreu nos *sites* das bibliotecas estrangeiras e foi realizada por meio do Protocolo de Coleta de Dados das Bibliotecas Estrangeiras, presente no Apêndice C. No que se refere ao OE3, a análise dos dados realizou-se com o retorno das quatorze instituições que compuseram a amostra através de onze gráficos e um quadro. Para preservar a identidade dos respondentes dos sistemas de bibliotecas das universidades do Nordeste brasileiro investigados, estes foram renomeados com um sistema alfanumérico constituído de quatro letras e um número, que vai de SIBI1 até SIBI14.

7 APRESENTAÇÃO E ANÁLISE DOS RESULTADOS

A presente seção é destinada à apresentação e análise dos resultados obtidos a partir da análise dos *sites* de três bibliotecas estrangeiras e do questionário aplicado para às coordenações e diretorias dos 14 sistemas de bibliotecas. É oportuno esclarecer que não houve a necessidade de análise da questão 1 (um) do questionário, pois se tratava apenas da identificação da instituição à qual o sistema de bibliotecas estava vinculado. Para melhor compreensão dos resultados, decidiu-se apresentar os SPGDC estruturados em dois eixos: o primeiro para expor os resultados referentes às bibliotecas estrangeiras e o segundo, por sua vez, o panorama das bibliotecas universitárias federais do Nordeste brasileiro.

7.1 Serviços e produtos para GDC em bibliotecas estrangeiras

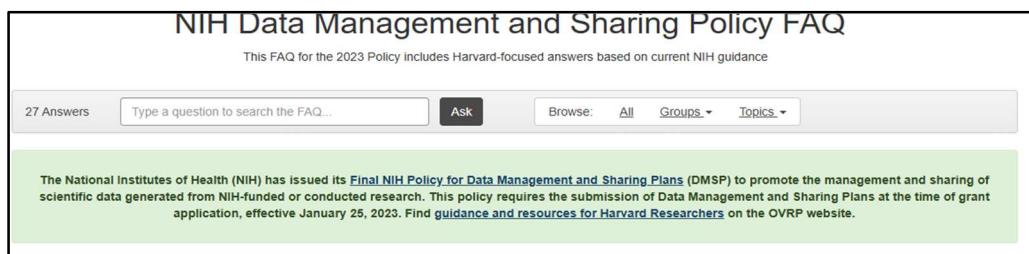
Os SPGDC ofertados pelas bibliotecas das Universidades de *Harvard*, *MIT* e *Cambridge* permitem entender como as recomendações da Ciência Aberta, orientada para gestão dos dados científicos, podem ser incorporadas ao cotidiano das bibliotecas universitárias. Essa prática proporciona informações precisas para que a comunidade acadêmica se adeque às mudanças ocorridas no ecossistema científico.

A seguir, serão apresentados alguns desses serviços e produtos em destaque nos *sites* dessas bibliotecas estrangeiras. A análise consistiu em investigar como o SPGDC é ofertado, através de políticas, treinamentos, cursos, oficinas, palestras, divulgação, orientações, manuais ou guias. Para melhor compreensão, decidiu-se por apresentar um extrato desses serviços e produtos.

Um dos itens investigados nos *sites* das bibliotecas estrangeiras foi a presença de política de gestão de dados científicos. No caso da Universidade de Harvard, apesar das bibliotecas não possuírem uma política para GDC, é importante destacar a Política de Preservação de dados do Repositório Harvard Dataverse, que é gerenciado pela Biblioteca e em conjunto com outras instâncias da Universidade. A Política prevê que os dados depositados no Harvard Dataverse serão preservados através de um modelo de rede distribuída. Esse modelo, segundo Márdero-Arellano (2012), é o mais indicado para garantir o acesso, compartilhamento, recuperação e preservação dos objetos digitais a longo prazo.

Vale destacar que, em alguns casos, são formulados documentos para orientar os pesquisadores a atenderem aos critérios das agências de fomento à pesquisa, as quais determinam o compartilhamento dos dados científicos de uma pesquisa científica. Um exemplo desse tipo de documento é o caso daquele que atende aos requisitos da *National Institutes of Health (NIH)*, uma das principais agências de fomento à pesquisa dos Estados Unidos. Nesse contexto, foi formulado um documento intitulado *NIH Data Management and Sharing Policy FAQ*¹¹² pelas Bibliotecas da Universidade de *Harvard*, com o objetivo de orientar os pesquisadores sobre a política de gerenciamento e compartilhamento de dados da *NIH*, conforme a Figura 7.

Figura 7 – FAQ sobre a Política do NIH



The screenshot shows the homepage of the NIH Data Management and Sharing Policy FAQ. The title is "NIH Data Management and Sharing Policy FAQ". Below the title, a subtext reads: "This FAQ for the 2023 Policy includes Harvard-focused answers based on current NIH guidance". There is a search bar with placeholder text "Type a question to search the FAQ...", a "Ask" button, and a "Browse" dropdown menu with options "All", "Groups", and "Topics". At the bottom of the page, a green box contains a note: "The National Institutes of Health (NIH) has issued its [Final NIH Policy for Data Management and Sharing Plans \(DMSP\)](#) to promote the management and sharing of scientific data generated from NIH-funded or conducted research. This policy requires the submission of Data Management and Sharing Plans at the time of grant application, effective January 25, 2023. Find [guidance and resources for Harvard Researchers](#) on the OVRP website."

Fonte: site das Bibliotecas da Universidade de *Harvard*.

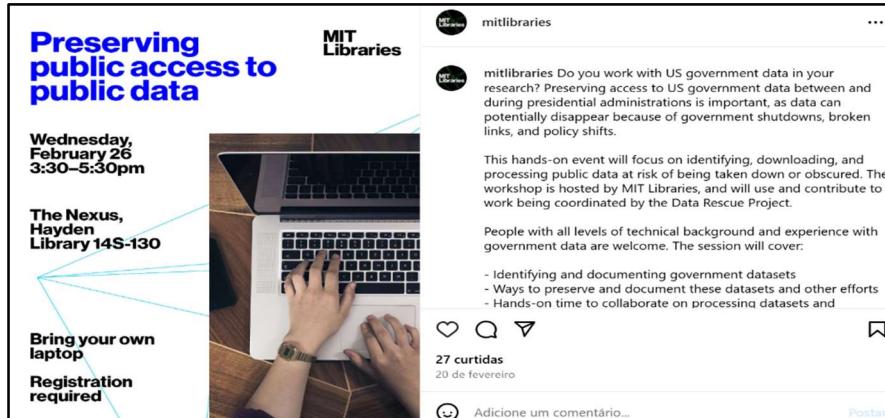
Outra iniciativa que merece destaque é a das Bibliotecas da Universidade de Cambridge, as quais criaram diversas políticas e diretrizes voltadas para o gerenciamento dos dados científicos, a exemplo da Política de Termos de Uso do seu Repositório de Dados (*Office of Scholarly Communication; Open Research Systems*, 2022), da Política do DOI no Repositório de Dados (Martinez-Garcia, 2023) e da Política de Preservação Digital da Biblioteca (Smith, 2021).

Em relação a divulgação de informações sobre GDC, observa-se que as bibliotecas recorrem a diferentes canais de comunicação, a exemplo de *sites*, redes sociais, e-mail institucionais. As bibliotecas do MIT utilizam o Instagram para a divulgação de eventos de gestão de dados promovidos pelo próprio setor. Em fevereiro de 2025, foi divulgado um *workshop* sobre o uso de dados governamentais na pesquisa científica, representada na Figura 8. O *workshop* consistia na aplicação de aulas práticas para a identificação, documentação e preservação de conjuntos de

¹¹² Ver informações em: <https://asklib.hms.harvard.edu/nih-dmfp>

dados governamentais que correm risco de serem removidos ou de ter o acesso restrinrido (*MIT Libraries*, 2025).

Figura 8 – Workshop dados governamentais na pesquisa científica



Fonte: Instagram das Bibliotecas do *MIT*¹¹³.

Já nas Bibliotecas da Universidade de *Cambridge*, o e-mail institucional é utilizado para envio de informações relevantes sobre acesso aberto, gestão de dados científicos, comunicação científica e tópicos de pesquisa aberta. A cada três meses, é enviado um boletim eletrônico para as pessoas que solicitaram cadastro na lista do e-mail. A Figura 9, a seguir, expõe o serviço.

Figura 9 – Boletim eletrônico



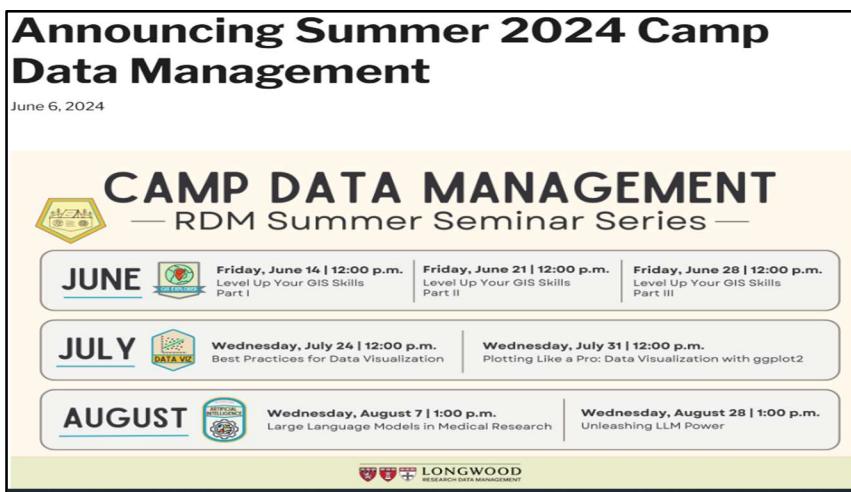
Fonte: site das Bibliotecas da Universidade de *Cambridge*¹¹⁴.

¹¹³ Ver informações em: <https://www.instagram.com/p/DGTTpifzT1t/>

¹¹⁴ Ver informações em: <https://www.data.cam.ac.uk/datanews>

Nas Bibliotecas de *Harvard*, identificou-se que o site das bibliotecas é um dos principais meios para divulgar *workshops*, eventos, serviços e produtos sobre GDC. Na Figura 10, apresenta-se um evento intitulado *Camp Data Management*¹¹⁵, realizado no verão de 2024 e organizado pela Biblioteca de *Logwood (Harvard)*. No evento, foram abordados *software* para visualização de dados, inteligência artificial aplicada para dados e outras temáticas relacionadas a GDC.

Figura 10 – Camp Data Management



Fonte: site das Bibliotecas da Universidade de *Harvard*.

No que se refere a realização de palestras sobre gestão de dados científicos para comunidade da universitária, as Bibliotecas de *Harvard* promovem periodicamente eventos organizados pelo próprio setor. Entre os meses de janeiro e maio de 2025, foram realizadas sete palestras. Nas Figuras 11 e 12, respectivamente podemos visualizar as palestras do mês de maio: uma sobre “Citação de dados” e outra sobre “Noções básicas de gerenciamento de dados científicos”. Essas palestras consistem em promover informações úteis para que os pesquisadores possam gerenciar os seus conjuntos de dados adequadamente (*Harvard*, 2025a).

¹¹⁵ Ver informações em: <https://datamanagement.hms.harvard.edu/news/announcing-summer-2024-camp-data-management>

Figura 11 – Evento sobre citação de dados



Fonte: site das Bibliotecas da Universidade de *Harvard*.

Figura 12 – Evento sobre gestão de dados



Fonte: site das Bibliotecas da Universidade de *Harvard*.

Já nas Bibliotecas do *MIT*, foi identificada a oferta de palestra com informações para que os pesquisadores da Instituição conheçam a nova Política sobre Planos de Gestão e Compartilhamento de Dados da *NIH*¹¹⁶. A palestra consistiu em apresentar as principais mudanças ocorridas na Política da *NIH*, bem como fornecer informações para auxiliar os pesquisadores na formulação de propostas para a obtenção de financiamento de suas pesquisas pela *NIH*, conforme a Figura 13.

¹¹⁶ Ver informações em: <https://grants.nih.gov/grants/guide/notice-files/NOT-OD-21-013.html>

Figura 13 – Palestra sobre a Política da NIH



Fonte: site das Bibliotecas do *MIT*¹¹⁷.

A presença de profissionais capacitados para a oferta de SPGDC foi constatada nas três bibliotecas, que disponibilizam para comunidade orientações, consultorias, cursos, treinamentos e criação de produtos para auxiliar os pesquisadores na gestão dos seus conjuntos de dados.

Nas Bibliotecas da Universidade de *Cambridge*¹¹⁸, a investigação identificou as seguintes funções:

- Coordenador e Gerente de dados científicos: responsáveis pelos serviços e treinamentos de GDC, pesquisa aberta, compartilhamento de dados, políticas de compartilhamento de dados das agências de fomento à pesquisa, PGD, orientação para o pesquisador escolher um repositório de dados e eventos em GDC.
- Gerente de sistemas de pesquisa aberta: responsável pela avaliação e integração do repositório institucional e fluxos de trabalho relacionados para depósitos de dados e publicações em outros sistemas, além de orientar os pesquisadores na formulação de documentação para dados, escolha de padrões de metadados e formatos de arquivos para dados complexos.

¹¹⁷ Ver informações em:

https://www.dropbox.com/scl/fi/jhxk7yglq2xlp8cs95hf/NIHdmsp_Slides/MIT.pdf?rlkey=q15omz9r1kzi2fywcc2cpmqou&e=1&dl=0

¹¹⁸ Ver informações em: <https://www.data.cam.ac.uk/research-data-management-team>

- Gerente de repositório: responsável pela orientação dos depósitos dos conjuntos de dados no repositório institucional e na escolha de outros repositórios (temáticos ou multidisciplinares) para proporcionar uma maior visibilidade dos conjuntos de dados.

Nas Bibliotecas do *MIT*¹¹⁹, a investigação identificou as seguintes funções:

- Diretor e Gerente de Serviços de Gerenciamento de Dados Científicos: responsável no suporte ao gerenciamento de dados do planejamento à publicação, orientação em relação aos requisitos das agências de fomento à pesquisa, orientação para formulação do PGD, criação de documentação para dados, preservação de dados e a seleção de repositórios (temáticos ou multidisciplinares) para depósito dos conjuntos de dados.
- Bibliotecários Especialistas em Dados: os Departamentos do *MIT* também possuem profissionais bibliotecários para prestar suporte específico na GDC nas áreas da Matemática, Química, Estatística, Engenharia Elétrica e Ciência da Computação.

Nas Bibliotecas da Universidade de *Harvard*¹²⁰, a investigação identificou as seguintes funções:

- Gerente de dados, Especialista em GDC, Especialista em conformidade de dados (orienta os pesquisadores sobre as exigências das agências de fomento à pesquisa e dos periódicos científicos), Bibliotecárias Especialistas em Dados em áreas específicas do conhecimento, curador de dados, Segurança de dados e Diretora Assistente em GDC.

Apesar das Bibliotecas da Universidade de *Harvard* possuírem profissionais capacitados para a demanda de GDC, a Universidade ainda mantém parceria com a *Research Data Management Librarian Academy (RDMLA)*¹²¹ como uma das alternativas de capacitação do seu corpo técnico, representada na Figura 14.

¹¹⁹ Ver informações em: <https://libguides.mit.edu/c.php?g=176063&p=1160988>

¹²⁰ Ver informações em: <https://datamanagement.hms.harvard.edu/about/who-we-are>

¹²¹ Ver informações em: <https://rdmla.github.io/>

Figura 14 – Research Data Management Librarian Data Academy



Fonte: site da *RDMLA*.

A *RDMLA* é uma plataforma que oferece diversos cursos focados na temática de dados científicos para capacitar bibliotecários e profissionais da informação, com a disponibilização de diversos cursos, como: Fundamentos da Gestão de Dados Científicos; Navegando pela cultura de dados científicos; Advogando e comercializando o valor do GDC em bibliotecas; Lançamento de serviços de dados em bibliotecas; Visão geral das ferramentas de análise e visualização de dados; Oferecendo treinamento em gerenciamento de dados: um guia para recursos do *DataONE*; Direitos autorais, licenciamento e privacidade de dados científicos e Curadoria e arquivamento de dados.

Para as bibliotecas que estão iniciando a capacitação dos seus profissionais, o estabelecimento de parceiras com outras instituições para a promoção de cursos em gestão de dados são uma alternativa. Todavia, é importante enfatizar que a capacitação em GDC é um processo contínuo e não estanque, por conta das diversas tipologias de dados existentes que, muitas vezes, requerem processos distintos em seu gerenciamento (Fearon *et al.*, 2013; Pinfield; Cox, 2014).

Sobre os Princípios *FAIR*, as Bibliotecas de *Harvard* promovem orientações aos seus usuários em relação ao seu uso. A Figura 15 expõe uma apresentação *online* realizada em 2023, intitulada “*Introduction to Data management and FAIR Data*¹²²”, com o objetivo de explorar os recursos do Repositório *Harvard Dataverse* aplicados aos Princípios *FAIR*; e a Figura 16 demonstra um passo a passo que foi disponibilizado pela biblioteca de como aplicar os Princípios *FAIR* para a pesquisa científica.

¹²² Ver informações em: <https://support.dataverse.harvard.edu/event/introduction-data-management-fair>

Figura 15 – Explicando os Princípios FAIR



Fonte: site das Bibliotecas da Universidade de *Harvard*.

Figura 16 - Aplicação dos Princípios FAIR

FAIR Data Principles: How can I apply them to my study?

John Obrycki | July 15, 2021

An NIEHS-funded administrative supplement to an existing project (R01 ES027825) led by Dr. Maitreyi Mazumdar (Boston Children's Hospital, Harvard Medical School) evaluated several approaches for applying FAIR principles. This data management-focused administrative supplement included a new collaboration with Julie Goldman, Research Data Services Librarian at Harvard Medical School's Countway Library. This collaboration identified several steps researchers can consider when developing data management plans for their studies.



Image credit: [LabTwin](#)

Fonte: site das Bibliotecas da Universidade de *Harvard*¹²³.

Já nas Bibliotecas da Universidade de *Cambridge*, conforme representado na Figura 17, foi ofertado um *workshop* para ensinar aos estudantes de pós-graduação como tornar os seus dados em FAIR, com espaço para discussão ao término do curso.

¹²³ Ver informações em: <https://datamanagement.hms.harvard.edu/news/fair-data-principles-how-can-i-apply-them-my-study>

Figura 17 – Workshop sobre os Princípios FAIR

University of Cambridge Training

Home Your bookings Timetable Courses Themes Venues

Mon 25 Mar 2019
14:30 - 16:30

Venue: Cambridge University Library, Milstein Room

Provided by: Office of Scholarly Communication

How FAIR is your research data?: a workshop (for researchers and postgraduate students in all disciplines)

Prerequisites New

Fonte: *site das Bibliotecas da Universidade de Cambridge*¹²⁴.

Compreender o ciclo de vida dos dados científicos é necessário para proceder com a curadoria digital e preservação dos dados. Na Figura 18, está exposto um curso sobre o ciclo de vida dos dados aplicado na pesquisa científica, ofertado pelas Bibliotecas da Universidade de Cambridge. O curso é dividido em módulos, como: armazenamento e *backup* de dados, melhores estratégias para organização de dados e compartilhamento de dados.

Figura 18 – Curso sobre o ciclo de vida dos dados

UNIVERSITY OF CAMBRIDGE CaRM Cambridge Research Methods

Modules Bookings Module Resources External Participants About

Research Data Management throughout the research life cycle (LT)

5th February 2025

Fonte: *site das Bibliotecas da Universidade de Cambridge*¹²⁵.

Nas Bibliotecas de *Harvard* e do *MIT*, é disponibilizado o *Open Science Framework*, ilustrado nas Figuras 19 e 20. A ferramenta possibilita o gerenciamento dos conjuntos de dados das pesquisas e projetos de pesquisa em todo o seu ciclo de vida, além de permitir a interação de pesquisadores que não fazem parte da pesquisa ou projeto para colaborarem com a sua construção.

¹²⁴ Ver informações em: <https://www.training.cam.ac.uk/event/2902978>

¹²⁵ Ver informações em: <https://www.researchmethods.group.cam.ac.uk/modules/research-data-management-throughout-the-research-life-cycle-lt/>

Figura 19 – Open Science Framework (Harvard University)

Open Science Framework

Use OSF to collaborate, manage, and share your documents, datasets, and research projects.

WHAT IS OSF?

Open Science Framework (OSF) is a free and open-source project management tool that makes it easy to collaborate throughout a project's lifecycle.

Fonte: *site* das Bibliotecas da Universidade de *Harvard*¹²⁶.

Figura 20 – Open Science Framework (MIT)

OSFINSTITUTIONS ▾
Search
Support
Donate
Sign Up

Service disruption event on Friday, May 23 that may have affected your OSF activity. Please see <https://www.cos.io/about/news/notice-of-osf-service-disruption> for more information



Massachusetts Institute of Technology

A research data service provided by the [MIT Libraries](#). Learn more about [MIT resources for data management](#). Please abide by the Institution's policy on [Privacy and Disclosure of Information](#).

Fonte: *site* das Biblioteca do *MIT*¹²⁷.

Na Biblioteca de Medicina de *Harvard*, também foi elaborado um Ciclo de Vida de Dados para contemplar dados biomédicos. O Instrumento foi elaborado com o objetivo de orientar os pesquisadores da área durante os processos de sua pesquisa, sendo formado pelas seguintes etapas: Planejar e Projetar, Coletar e Criar, Analisar e Colaborar, Avaliar e Arquivar, Compartilhar e Disseminar e Publicar e Reutilizar, conforme a Figura 21.

¹²⁶ Ver informações em: <https://library.harvard.edu/services-tools/open-science-framework>

¹²⁷ Ver informações em: <https://osf.io/institutions/mit>

Figura 21 – Ciclo de Vida dos Dados na área Biomédica



Fonte: site das Bibliotecas da Universidade de *Harvard*¹²⁸.

O ciclo de vida de um dado é influenciado fortemente pela sua heterogeneidade, isso quer dizer que dados gerados em distintas disciplinas ou áreas do conhecimento requerem processos diferentes em sua coleta, descrição, armazenagem, preservação, acesso, compartilhamento e descarte. O objetivo é proporcionar tratamentos adequados para abranger as múltiplas tipologias de dados e os diversos propósitos dos pesquisadores e instituições (Sales, 2014; Plale; Kouper, 2017).

Acerca dos planos de gestão de dados, as Bibliotecas do *MIT*, *Harvard* e *Cambridge* disponibilizam uma série de informações em seus *sites* para auxiliar os pesquisadores no preenchimento de seus PGD, além de oferecer serviços personalizados para preenchimento e correção do PGD. Os pesquisadores também podem utilizar as suas credenciais institucionais para acessar o *DMPtool* (*Harvard* e *MIT*)¹²⁹ e o *DMPonline* (*Cambridge*)¹³⁰ para o preenchimento do seu PGD em formato digital, além de ter a sua disposição modelos de PGD já preenchidos nessas ferramentas, conforme as Figuras 22, 23 e 24.

¹²⁸ Ver informações em: <https://datamanagement.hms.harvard.edu/plan-design/biomedical-data-lifecycle>

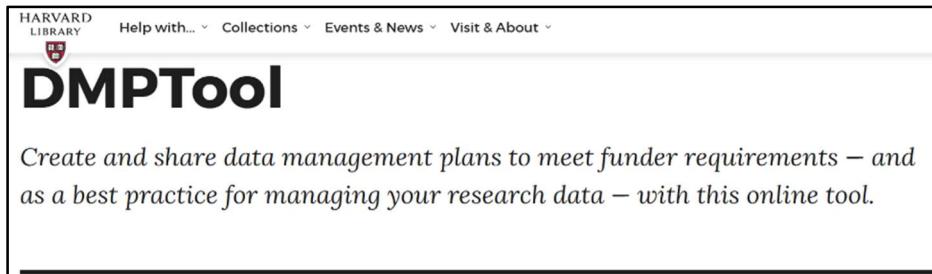
¹²⁹ Ver informações em:

<https://library.harvard.edu/services-tools/dmptool>

<https://libraries.mit.edu/data-management/plan/write/>

¹³⁰ Ver informações em: <https://www.data.cam.ac.uk/data-management-guide/creating-your-data/data-management-plan>

Figura 22 – DMPTool (Harvard University)



Fonte: *site* das Bibliotecas da universidade de *Harvard*.

Figura 23 – DMPTool (MIT)



Fonte: *site* das Bibliotecas do *MIT*.

Figura 24 – DMPonline (University of Cambridge)



Fonte: *site* das Bibliotecas da Universidade de *Cambridge*.

Nas Bibliotecas da Universidade de *Cambridge*, também são promovidos cursos para capacitação dos profissionais bibliotecários no preenchimento do PGD, conforme a Figura 25. Nessa perspectiva, compreende-se que a capacitação dos responsáveis pelo serviço antecede a sua criação, corroborando com o que é preconizado por Fearon *et al.* (2013).

Figura 25 – Curso para bibliotecários sobre PGD

The screenshot shows a course listing on the University of Cambridge Training website. The course is titled "Managing Data Management: Getting Started with Data Management Plan Support (for librarians)". It is scheduled for Thursday, June 7, 2018, from 14:30 to 16:30, in the Milstein Room at the Cambridge University Library. The course is provided by the Office of Scholarly Communication. A "New" tag is visible next to the course title.

Fonte: *site* das Bibliotecas da Universidade de *Cambridge*¹³¹.

Os dados pessoais e sensíveis em pesquisa são um aspecto que deve ser observado, quando discutimos a gestão de dados científicos, pois o compartilhamento deles pode transgredir as legislações que orientam a sua anonimização. Entre as estratégias empregadas pelas bibliotecas estrangeiras para conscientizar os pesquisadores, identificaram-se eventos, elaboração de manuais e orientações por meio de FAQ. As Bibliotecas da Universidade de *Cambridge*, por exemplo, realizaram em maio de 2025 o “Fórum de Campeões de Dados” com a temática “Trabalhando com dados sensíveis”, conforme a Figura 26. Neste evento, são compartilhadas as boas práticas dos pesquisadores considerados “Campeões de Dados”.

Figura 26 – Evento sobre dados sensíveis

The screenshot shows the "Data Champions Forum May 2025" event page on the University of Cambridge Training website. The event is scheduled for Thursday, May 29, 2025, from 12:00 to 14:00, in Room GR06/07 at the Faculty of English. The event is provided by the Office of Scholarly Communication. A "New" tag is visible next to the event title. The "Booking" section indicates that places are available and provides a "Book a place" button. The "Description" section provides details about the forum, including its theme ("Working with Sensitive Data"), venue, lunch provision, and booking instructions.

Fonte: *site* das Bibliotecas da Universidade de *Cambridge*¹³².

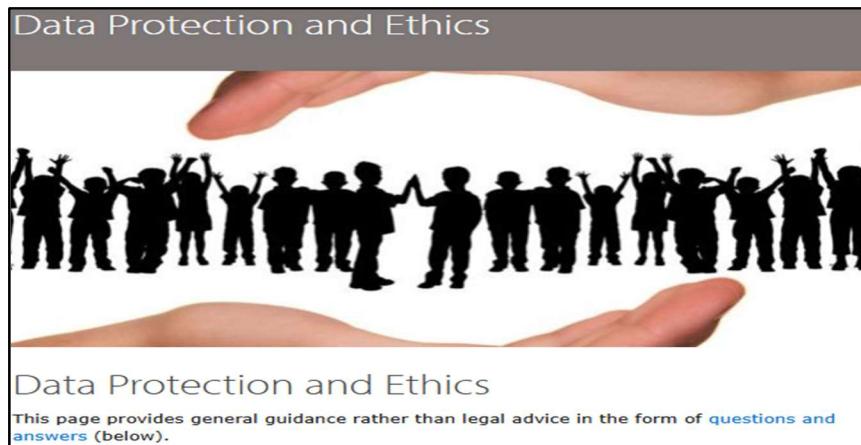
Como forma de orientar sua comunidade científica, as Bibliotecas da Universidade de *Cambridge* e do *MIT* disponibilizam em seus *sites* diversas informações em relação ao uso, compartilhamento e armazenamento de dados

¹³¹ Ver informações em: <https://www.training.cam.ac.uk/event/2536948>

¹³² Ver informações em: <https://www.training.cam.ac.uk/event/5753983>

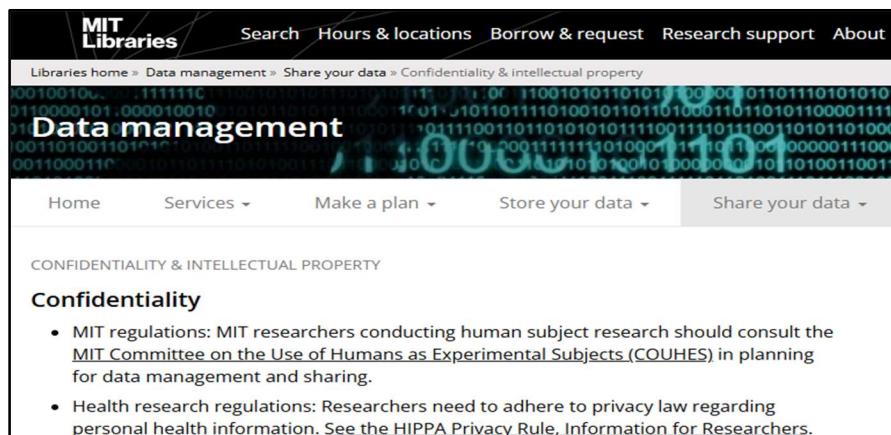
pessoais e sensíveis. Entre essas informações, podemos citar: FAQ, guias, políticas e recomendações, conforme representado nas Figuras 27 e 28.

Figura 27 – FAQ sobre dados pessoais e sensíveis



Fonte: *site das Bibliotecas da Universidade de Cambridge*¹³³.

Figura 28 - Informações sobre dados pessoais e sensíveis



Fonte: *site das Bibliotecas do MIT*¹³⁴.

A adoção de medidas de proteção dos conjuntos de dados é uma prática importante que visa garantir que os dados não sejam acessados por pessoas indevidas. As Bibliotecas da Universidade de *Harvard*, por exemplo, orientam o uso de determinadas práticas para garantir a segurança dos dados pessoais e sensíveis coletados pelos seus pesquisadores, como (*Harvard University*, 2025d):

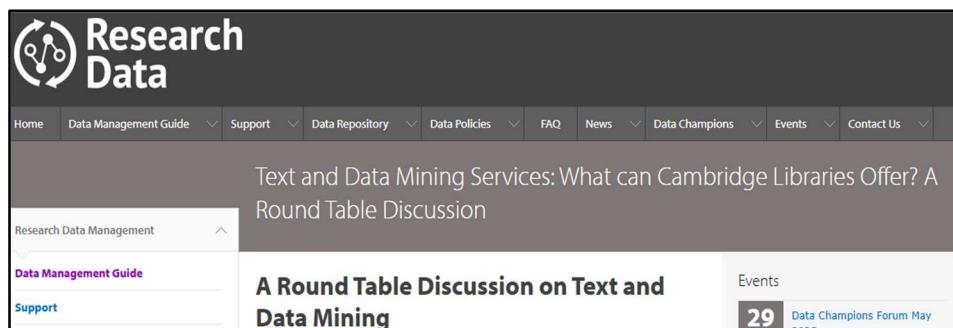
¹³³ Ver informações em: <https://www.data.cam.ac.uk/sensitive-data>

¹³⁴ Ver informações em: <https://libraries.mit.edu/data-management/share/confidentiality/>

- Segurança de rede: evitar o armazenamento de dados na internet e, caso não seja possível, considerar o uso de *firewall*;
- Segurança física: aplicar restrições para uso de computadores com o armazenamento de dados pessoais e sensíveis e utilizar apenas pessoas de confiança para realizar manutenção em seu computador;
- Sistemas e arquivos de computador: utilizar programas de antivírus e manter atualizados, evitar o uso do e-mail para envio de arquivos e proteger o acesso ao computador e arquivos com senhas;
- Criptografia: utilizar ferramentas de criptografia para proteger os arquivos.

Os serviços e produtos para orientar os pesquisadores a encontrar dados úteis para suas pesquisas foram constatados em duas universidades. Nas Bibliotecas da Universidade de Cambridge, podemos identificar, na Figura 29, um evento em formato de mesa redonda organizada para discutir o que a comunidade da instituição poderia esperar das bibliotecas no suporte à mineração de texto e dados em pesquisa científica. Esse tipo de evento é uma excelente oportunidade para a biblioteca conhecer as necessidades da sua comunidade em torno de um serviço específico.

Figura 29 – Mesa redonda sobre mineração de textos e dados



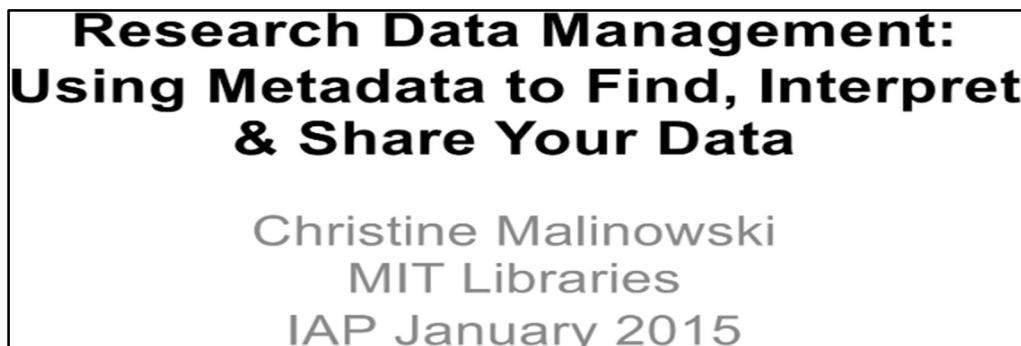
Fonte: site das Bibliotecas da Universidade de Cambridge¹³⁵.

Na Biblioteca do *MIT*, foi realizado um *workshop* (Figura 30) com o objetivo de orientar os pesquisadores sobre a importância da aplicação dos metadados para encontrar, interpretar e compartilhar dados. Nesse sentido, o *workshop* informa que metadados podem fornecer informações precisas sobre: os processos usados para a criação dos dados; o *software* necessário para a leitura dos dados; quais as restrições aplicadas para o compartilhamento dos dados; sob qual contexto os dados foram

¹³⁵ Ver informações em: <https://www.data.cam.ac.uk/text-and-data-mining-services-what-can-cambridge-libraries-offer-round-table-discussion>

criados; a possibilidade de interpretação dos dados e como os dados devem ser citados.

Figura 30 – Metadados para encontrar, interpretar e compartilhar dados



Fonte: *site das Bibliotecas do MIT*¹³⁶.

Sobre repositórios institucionais de dados científicos, as universidades de Cambridge e Harvard disponibilizam serviços de repositórios de dados científicos com infraestrutura construída nas próprias instituições. Os repositórios são chamados de *Apollo* (Universidade de *Cambridge*) e *Harvard Dataverse* (Universidade de *Harvard*), conforme as Figuras 31 e 32.

O *Apollo* é gerenciado pela *Open Research Systems* (ORS). A equipe é subordinada ao Sistema de Bibliotecas da Universidade de *Cambridge* e é responsável pelo desenvolvimento técnico do Repositório e por integrações com outras plataformas da Universidade, assim como a manutenção, atualização do software e atendimento ao usuário.

O *Harvard Dataverse* é gerenciado pela Equipe da Biblioteca de *Harvard* juntamente com o Departamento de Tecnologia da Informação da Universidade de Harvard. Apesar de ser um serviço institucional, o Repositório também recebe conjuntos de dados de pesquisadores de outras instituições que possuem parcerias mantidas com a *Harvard University*, ampliando, assim, a sua oferta de dados científicos.

¹³⁶ Ver informações em: https://ocw.mit.edu/courses/res-str-002-data-management-spring-2016/resources/mitres_str_002s16_usngmeta/

Figura 31 - Repositório de Dados Científico (Apollo)

Apollo is the institutional repository of the University of Cambridge, managed by the [Open Research Systems](#) team based in Cambridge University Library. The Repository is committed to store and preserve the University's research outputs. Research outputs can include, but are not limited to, publications, conference proceedings, book chapters, monographs, theses, various forms of research data (video recordings, spreadsheets, computational scripts, code, images etc.), presentations and others.

Fonte: [site do Repositório de Dados Científicos da Universidade de Cambridge](https://www.repository.cam.ac.uk/home)¹³⁷.

Figura 32 - Repositório de Dados Científico (Harvard Dataverse)

Deposit and share your data. Get academic credit.

Organize datasets and gather metrics in your own repository.

Publishing your data is easy on Harvard Dataverse!

Harvard Dataverse is a repository for research data. Deposit data and code here.

A dataverse is a container for all your datasets, files, and metadata.

Learn about getting started creating your own dataverse repository here.

Add a dataset +

Add a dataverse +

Getting started

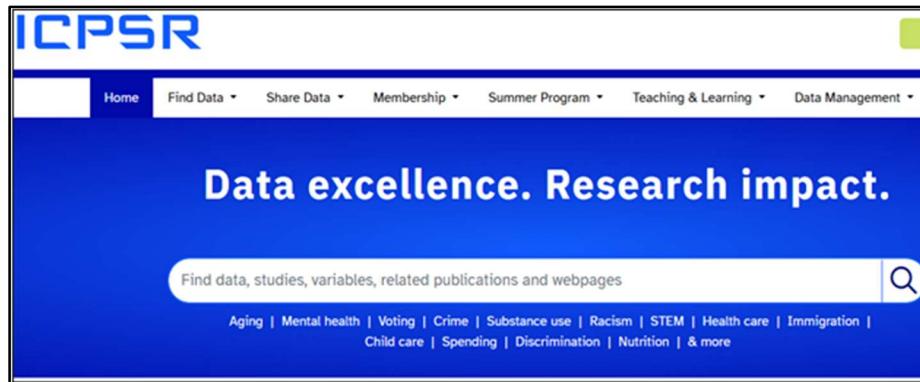
Fonte: [site do Repositório de Dados científicos da Universidade de Harvard](https://data.harvard.edu/)¹³⁸.

Já o *MIT* optou por estabelecer parcerias com outras instituições para disponibilizar o serviço de repositório de dados científicos para os membros da sua instituição, conforme as Figuras 33 e 34. Nesse sentido, são disponibilizados *The Inter-university Consortium for Political and Social Research (ICPSR)* especializado em Ciências Sociais e comportamental e o *Qualitative Data Repository (QDR)*, especializado nas áreas de ciências sociais, saúde, direito, educação e humanidades.

¹³⁷ Ver informações em: <https://www.repository.cam.ac.uk/home>

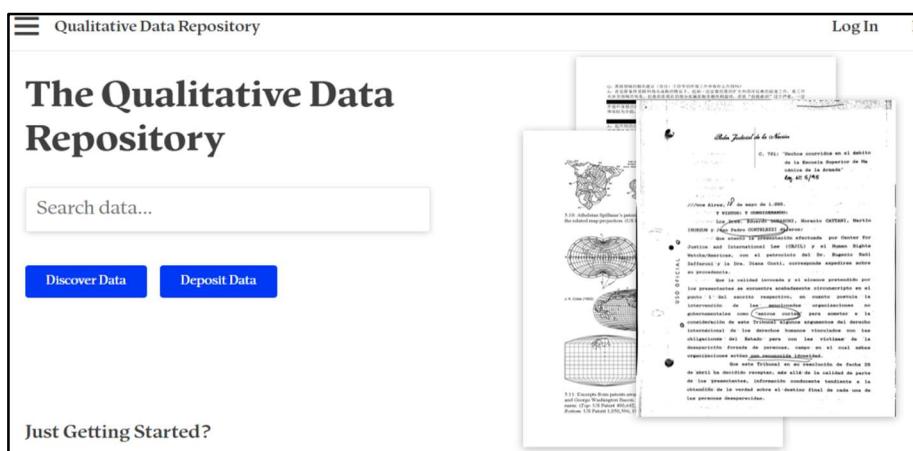
¹³⁸ Ver informações em: <https://data.harvard.edu/>

Figura 33 - Repositório de Dados Científico da ICPSR



Fonte: *site* do Repositório de Dados Científicos da Universidade de Michigan¹³⁹.

Figura 34 - Repositório de Dados Científico da QDR



Fonte: *site* do Repositório de Dados Científicos da Universidade de Syracuse¹⁴⁰.

Segundo Silva (2019), os repositórios de dados configuram-se atualmente como uma das principais fontes de informações para pesquisa de dados científicos, além de proporcionar a organização, integridade, tratamento, compartilhamento e preservação desses dados. Sobretudo, seus atributos os capacitam a auxiliar os pesquisadores no cumprimento dos requisitos das agências de fomento à pesquisa e periódicos científicos.

A indicação de repositório de dados que melhor atenda às necessidades do pesquisador é uma etapa importante para a gestão de dados científicos. De acordo com Silva (2016), é fundamental conhecer as características tecnológicas do repositório e verificar se estas poderão proporcionar vantagens ou desvantagens no

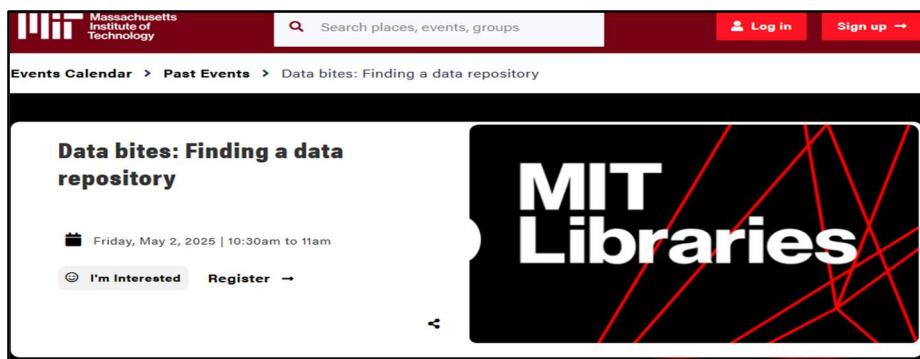
¹³⁹ Ver informações em: <https://www.icpsr.umich.edu/web/pages/index.html>

¹⁴⁰ Ver informações em: <https://qdr.syr.edu/>

compartilhamento, preservação e visibilidade dos dados. Aspectos como o uso de DOI, política de preservação, política de compartilhamento, tipologia do repositório (institucional, temático ou multidisciplinar) e interoperabilidade, destacam-se como fundamentais para uma escolha.

Seguindo os pré-requisitos citados anteriormente, nas bibliotecas do *MIT* são ofertados *workshops* com orientações para a comunidade da instituição identificar as principais características dos repositórios de dados que estão em conformidade com as recomendações do *MIT*, conforme Figura 35.

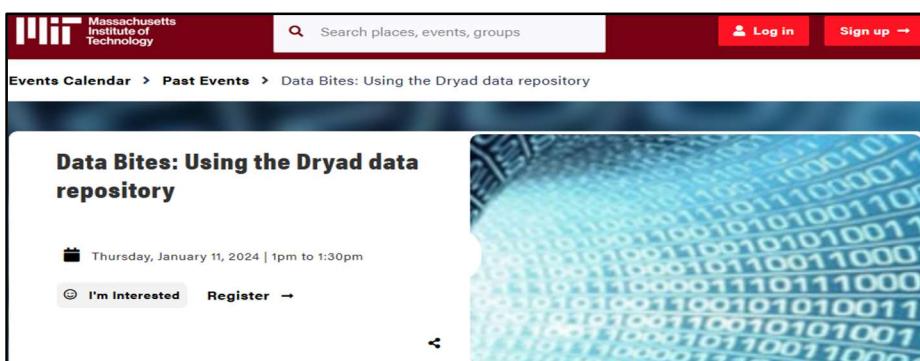
Figura 35 – Curso para encontrar repositórios de dados científicos



Fonte: *site* das Bibliotecas do *MIT*¹⁴¹.

Na Figura 36, é oferecido um *workshop* para os pesquisadores, que são financiados pelo *National Institutes of Health (NIH)*, encontrarem um repositório de dados que atenda às Políticas de gerenciamento e compartilhamento de dados do *NIH*. Nesse contexto, o repositório indicado pelo *workshop* é o *Dryad*.

Figura 36 – Workshop depositando no Repositório Dryad



Fonte: *site* das Bibliotecas do *MIT*¹⁴².

¹⁴¹ Ver informações em: <https://calendar.mit.edu/event/data-bites-finding-a-data-repository-7667>

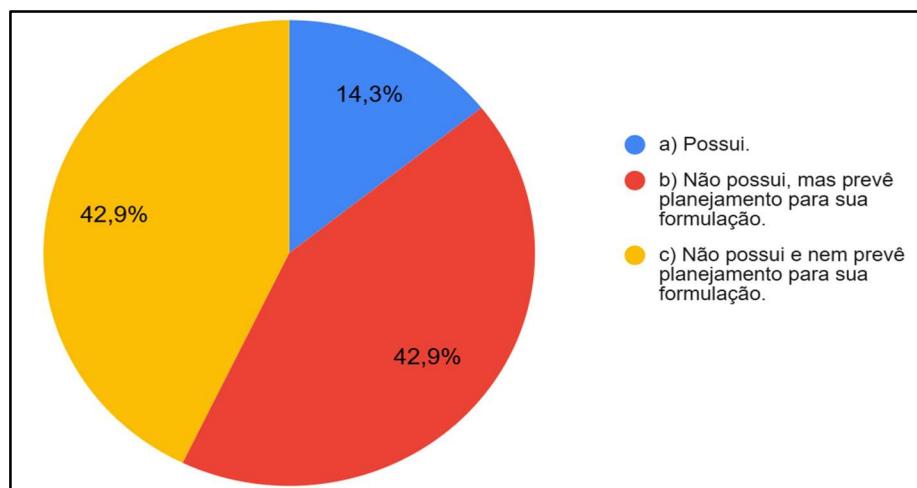
¹⁴² Ver informações em: https://calendar.mit.edu/event/data_bites_dryad_iap2024

As bibliotecas do *MIT* também disponibilizam uma Tabela¹⁴³ com características e funcionalidades de vários repositórios de dados (*Dataverse*, *Zenodo*, *Dryad* e *OSF*), para que os pesquisadores constatem qual repositório se adequa mais às suas necessidades.

7.2 Identificação de produtos e serviços nos sistemas de bibliotecas das universidades federais do Nordeste brasileiro

Ao serem questionados se **o sistema de bibliotecas possui política/norma implementada para gestão dos dados científicos dos pesquisadores da universidade**, constatou-se que um número diminuto de respondentes afirmou que o possui, representado por dois (14,3%) sistemas de bibliotecas. As opções “não possui, mas prevê planejamento para a sua formulação” e “não possui e nem prevê planejamento para a formulação de política/norma” foram respondidas por seis (42,9%) sistemas de bibliotecas para cada uma das opções (Gráfico 1).

Gráfico 1 - Política/Norma para gestão de dados científicos



Fonte: elaboração própria.

Entre os sistemas de bibliotecas que responderam positivamente para o item “Possui”, apenas o SIBI¹⁴ enviou o *link* para o acesso à política, conforme solicitado no item 2.1 do questionário. Contudo o documento fornecido é referente a Política do

¹⁴³ Ver informações em: <https://libraries.mit.edu/data-management/share/find-repository/>

Repositório Institucional de Publicações Científicas da instituição, não possuindo diretrizes para GDC. Ao término da análise constatou-se que nenhum sistema de bibliotecas possui o serviço e que oito (57,1%) não o possuem o serviço e nem preveem planejamento para sua formulação.

Os resultados interpretados de acordo com as respostas obtidas estão em desacordo com o que foi verificado na literatura científica, que preconiza que a política ou norma é um dos principais instrumentos que institucionaliza a gestão de dados científicos e orienta a tomada de decisões na instituição.

Para Jones, Pryor e White (2013), as políticas garantem que os serviços e produtos sejam eficientemente coordenados e alinhados com políticas institucionais e nacionais. Silveira *et al.* (2021) corroboram com os autores ao afirmar que as instituições podem incentivar a promoção de práticas da Ciência Aberta com a formulação de políticas, a partir de dispositivos como: regulamentos, portarias e instruções normativas, para estabelecer diretrizes para proceder com a criação de um serviço ou produto para gerenciar dados.

A inexistência de políticas ou normas para GDC evidenciada nesses sistemas de bibliotecas pode ser uma consequência da ausência de uma política pública nacional voltada para essa questão. Apesar dos avanços alcançados nos últimos anos, o Brasil ainda não estabeleceu uma política de referência para gerir os dados científicos que são produzidos no país.

Como exemplo desses avanços podemos citar, o Decreto nº 8.777 que foi a primeira política pública a abordar diretrizes para à abertura dos dados que são produzidos pelas IES, por meio de mecanismos, como a obrigatoriedade da formulação do Plano de Dados Abertos (PDA), assim como, os Planos de Ação Nacional para o Governo Aberto, especificamente o 4º, 5º e 6º (em andamento), que contêm elementos que contribuem para disseminação e preservação dos dados científicos. Entretanto esses documentos não se configuram em políticas nacionais com diretrizes específicas e abrangentes para o gerenciamento dos dados científicos que são produzidos no Brasil.

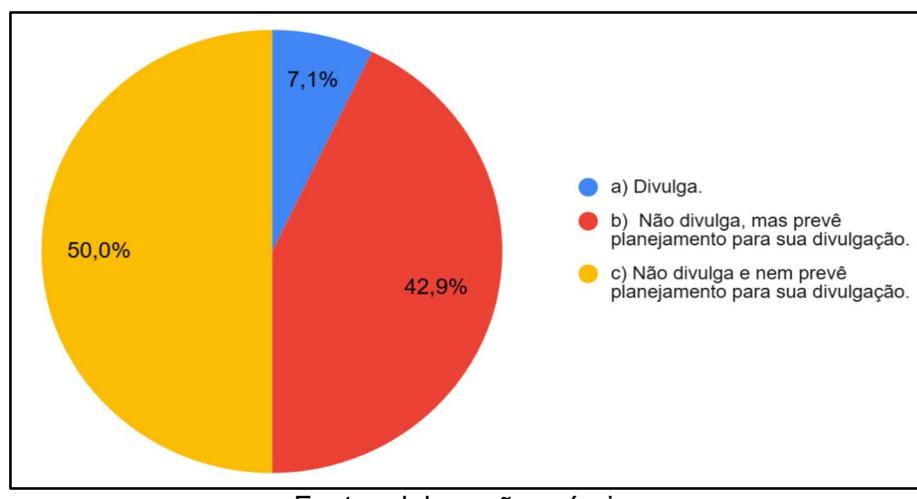
Nessa perspectiva, com a ausência de uma política pública para norteia à gestão dos dados científicos que são gerados nas IES, essas instituições podem fazer uso dos dispositivos contidos no Decreto nº 8.777, Lei nº 12.527 e nos 4º, 5º e 6º Planos de Ação Nacional para o Governo Aberto, como base para a construção de

suas políticas para GDC. Importante ressaltar que a construção dessas políticas visa não somente promover a governança dos dados científicos nas IES, mas, também, atender aos princípios contidos na Constituição Federal de 1988, que no seu Artigo 5º, inciso XXXIII, estabelece o direito ao acesso as informações dos órgãos públicos aos cidadãos.

Pretendeu-se consultar se **o sistema de bibliotecas divulga informações sobre gestão de dados científicos através do seu site, das suas redes sociais ou e-mail institucional**.

Os resultados revelaram que apenas um (7,1%) sistema de bibliotecas informou que realiza o serviço, porém o *link* fornecido pelo respondente na resposta 3.1 do questionário para comprovação do serviço, remete à página do Repositório Institucional de Publicações Científicas da instituição, o que não comprovou a efetividade do referido serviço. Para seis (42,9%) respondentes, o serviço se encontra em vias de planejamento e sete (50,0%) sistemas de bibliotecas optaram pela resposta não divulgam e nem preveem planejamento para sua divulgação (Gráfico 2).

Gráfico 2 - Divulgação de informações sobre GDC



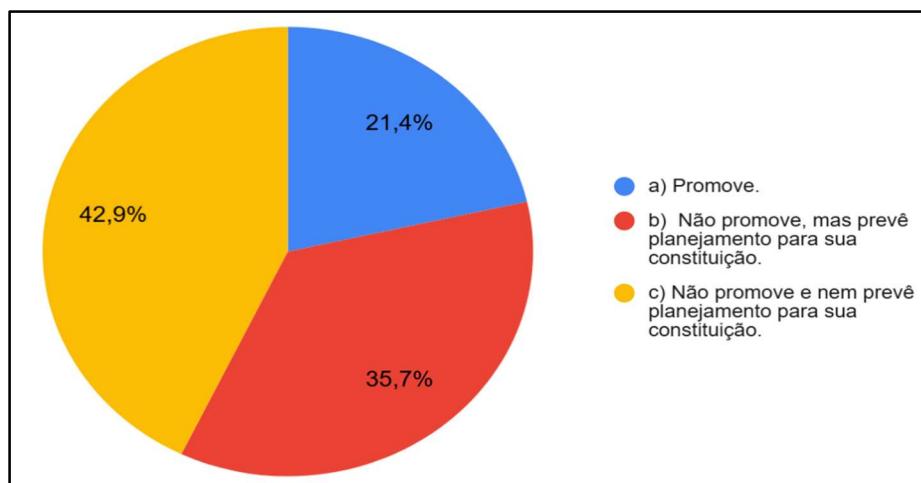
Ao término da análise, constatou-se que nenhum sistema de bibliotecas realiza o serviço e que oito (57,1%) não possuem o serviço e nem preveem planejamento para a sua formulação, apontando, assim, uma grande lacuna nesta oferta.

O fato revela-se preocupante, por compreender que essas ferramentas de comunicação representam uma grande possibilidade de alcance para informar e conscientizar os pesquisadores da relevância da GDC. A divulgação de informações

sobre GDC em *sites* e redes sociais da instituição constitui excelente estratégia para estabelecer interação com a comunidade acadêmica e compreender as suas necessidades, além de produzir métricas dos serviços oferecidos pela instituição (Costa; Araújo; Paixão, 2024).

Quando inquiridos se **o sistema de bibliotecas promove ações de sensibilização sobre a importância do gerenciamento dos dados científicos para comunidade da universidade através de palestras**, três (21,4%) respondentes afirmaram possuir o serviço, cinco (35,7%) sistemas informaram não promovem, mas preveem planejamento para sua constituição e seis (42,9%) optaram por marcar não promovem e nem preveem planejamento para sua constituição. Informações dispostas no Gráfico 3.

Gráfico 3 - Palestras sobre GDC



Fonte: elaboração própria.

Os três sistemas de bibliotecas que responderam positivamente ao item “Promove” enviaram informações sobre o referido serviço.

O SIBI8 informou que “As palestras versam sobre a definição de dado, o que são dados de pesquisa e de que maneira gerenciá-los. Além disso, promovemos capacitações sobre o Plano de Gestão de Dados, Princípios FAIR e CARE, e outros temas correlatos relacionados aos dados e à Ciência Aberta”.

O SIBI11 informou que são realizados “Treinamentos sobre a plataforma de Gestão de TCC e Reuniões sobre a importância e mediação do processo de gestão de dados através dos produtos de conclusão de curso gerados pelos concluintes”.

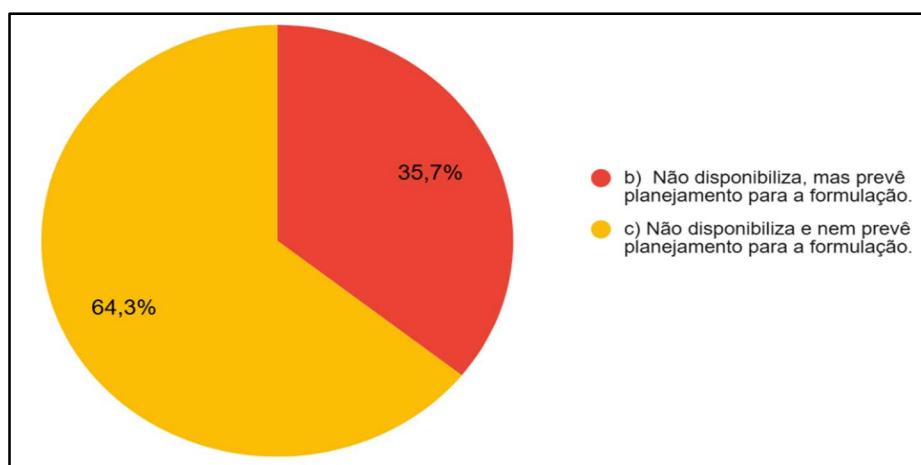
O SIBI12 informou que as palestras foram realizadas através de “Lives do Projeto Registros de Informação” e do “III ENCONTRO BAIANO DE BIBLIOTECAS UNIVERSITÁRIAS PÚBLICAS - EBBUP”.

Ao término da análise, constatou-se que as ações descritas pelos sistemas de bibliotecas indicam temáticas que convergem com a GDC, a exemplo do “plano de gestão de dados”, “Princípios FAIR e CARE” e “processo de gestão de dados”. Embora o SIBI12 não tenha explicitado as temáticas abordadas durante as ações.

Nessa perspectiva, palestras e eventos temáticos são excelentes estratégias para estabelecer os primeiros contatos entre as bibliotecas e a comunidade da universidade. As bibliotecas podem desenvolver palestras específicas para cada tipo de público, apenas com conteúdos introdutórios ou mais avançados, assim como a criação de palestras interativas, separando um momento com perguntas ao final de cada apresentação.

No que concerne se **o sistema de bibliotecas disponibiliza suporte financeiro para participação de bibliotecários(as) em cursos de capacitação em gestão de dados científicos**, constatou-se que quatorze (100%) sistemas de bibliotecas não dispõem de recursos financeiros, para capacitação dos profissionais, conforme o Gráfico 4.

Gráfico 4 – Verba para capacitação de bibliotecários em GDC



Fonte: elaboração própria.

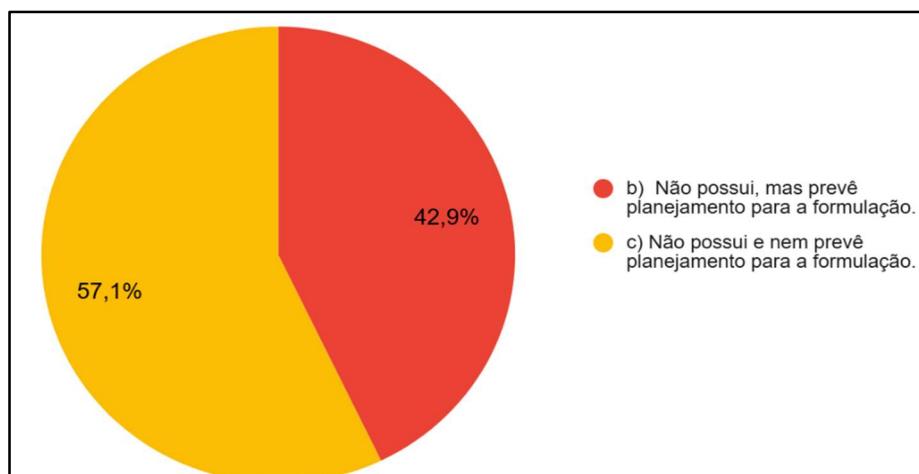
Embora nenhuma instituição investigada disponha de investimento financeiro para capacitação dos bibliotecários em GDC, cinco (35,7%) sistemas de bibliotecas preveem planejamento para o seu desenvolvimento, enquanto nove (64,3%) instituições não preveem planejamento para sua constituição. Os números

demonstram uma lacuna na oferta desse serviço direcionado aos profissionais bibliotecários, o que pode impactar na promoção de serviços de mais qualidade e na formação de profissionais mais capacitados para identificar as múltiplas necessidades de sua comunidade. Sales *et al.* (2019) afirmam que, para promoverem serviços voltados para GDC nas perspectivas tecnológica, legais e éticas, os profissionais bibliotecários necessitam de capacitação para compreender a complexidade em torno do tratamento de dados científicos, além disso, esses profissionais capacitados podem participar ativamente do processo de geração de dados durante a pesquisa com orientações específicas aos pesquisadores.

Para as bibliotecas que não possuem suporte financeiro para a capacitação dos seus profissionais é possível recorrer aos cursos gratuitos. As plataformas da *Open Plato* e da Fiocruz disponibilizam diversos cursos de capacitação em GDC divididos em módulos que abordam diversas informações sobre a temática.

Quando questionados se **o sistema de bibliotecas possui serviço de orientação sobre o uso dos Princípios FAIR**, nenhum dos respondentes informou possuir o serviço, porém, seis (42,9%) unidades, apesar de não possuírem o serviço, já possuem planejamento para a sua concepção e enquanto oito (57,1%) não o possuem e nem preveem planejamento para a formulação.

Gráfico 5 - Orientação sobre Princípios FAIR



Fonte: elaboração própria.

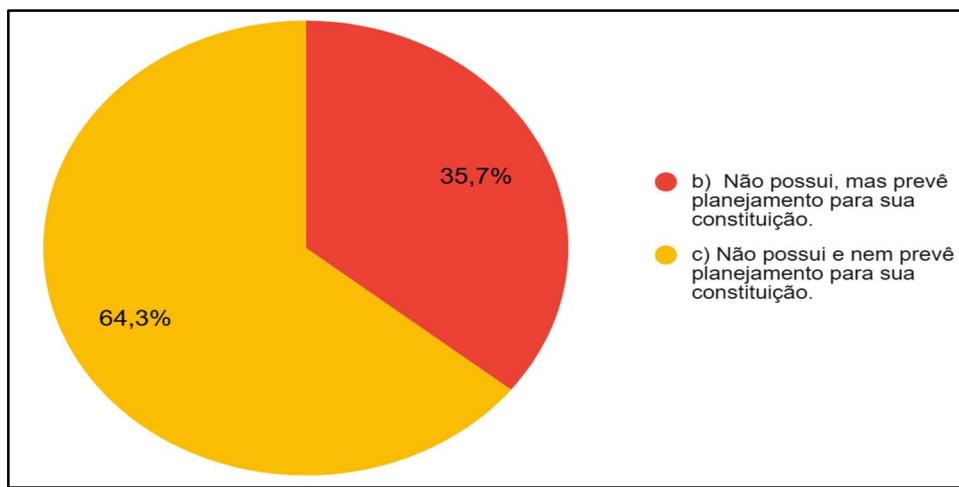
Esses resultados indicam a necessidade da concepção desse serviço, pois, por meio do uso dos Princípios FAIR, os pesquisadores podem gerenciar os seus dados

científicos de forma mais adequada e proporcionar maior visibilidade e reutilização dos dados de sua pesquisa.

A aplicação dos Princípios *FAIR* serve como um guia orientador na promoção do gerenciamento de dados. Seus mecanismos são os mais adequados para tornar os dados científicos em encontráveis, acessíveis, interoperáveis e reutilizáveis. Nesse cenário os Princípios *FAIR* fornecem escolhas para que pesquisadores ou gerenciadores de infraestruturas eletrônicas superem obstáculos para estabelecer a descoberta de dados científicos por máquinas (Wilkinson *et al.*, 2016).

Quanto à pergunta se **o sistema de bibliotecas possui serviço de orientação sobre o funcionamento do Ciclo de Vida dos Dados na gestão de dados científicos**, cinco (35,7%) respondentes informaram que não possuem, mas preveem planejamento para a sua constituição, enquanto nove (64,3%) marcaram a opção não possuem e nem preveem planejamento para sua constituição, segundo os dados presentes no Gráfico 6.

Gráfico 6 – Orientação sobre o ciclo de vida dos dados



Fonte: elaboração própria.

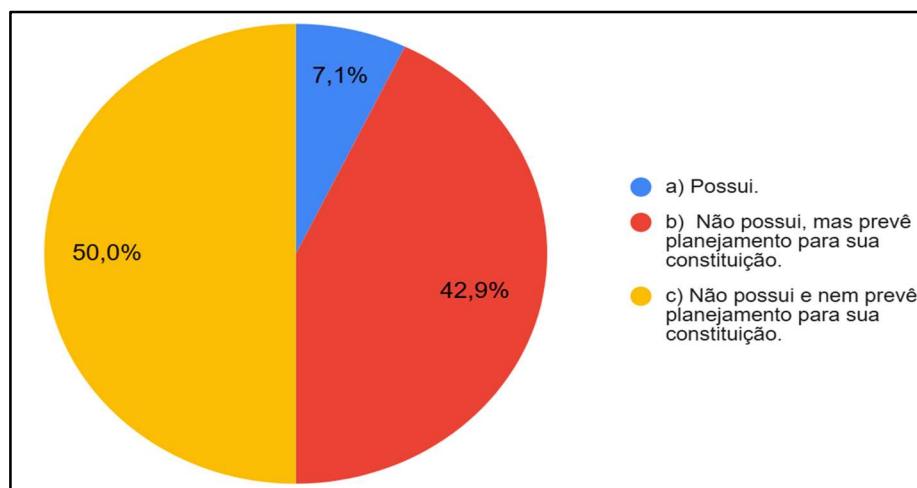
Nessa conjuntura, as respostas anteriores expõem a inexistência do referido serviço nos sistemas de bibliotecas, configurando um prejuízo significativo na orientação aos pesquisadores sobre o funcionamento dos diferentes ciclos de vida dos dados. Nas palavras de Sayão e Sales (2020), é justamente por meio do conhecimento sobre os mecanismos do ciclo de vida que será realizada a correta curadoria dos dados. A aquisição desse conhecimento se faz necessária, pois existe

uma grande variedade de tipologias de dados, necessitando, às vezes, de abordagens distintas em seu tratamento.

A criação de serviços voltados para orientar os pesquisadores sobre o ciclo de vida dos dados também é fundamental para que o processo de curadoria não seja afetado negativamente e prejudique o reuso dos dados.

No que concerne ao questionamento se **o sistema de bibliotecas possui serviço de orientação para o preenchimento do Plano de Gestão de Dados (PGD) das pesquisas desenvolvidas na universidade**, apenas um (7,1%) sistema de bibliotecas afirmou possuir o serviço. Entretanto a unidade não enviou a comprovação do referido serviço conforme solicitado na pergunta subsequente do questionário. Outras seis (42,9%) unidades não possuem, mas preveem planejamento para a sua constituição e sete (50,0%) unidades não possuem e nem preveem planejamento para a sua constituição, conforme o Gráfico 7.

Gráfico 7 - Orientação para o preenchimento do Plano de Gestão de Dados



Fonte: elaboração própria.

Ao término da análise das respostas cedidas pelos respondentes dos sistemas de bibliotecas, constatou-se que não existe o referido serviço em nenhuma das unidades, assim o quantitativo de bibliotecas que não possuem o serviço e nem preveem planejamento para sua formulação subiu para oito (57,1%).

O número diminuto de sistemas de bibliotecas de universidades federais do Nordeste brasileiro que possuem PGD pode resultar em barreiras para os pesquisadores, à medida que as agências de fomento à pesquisa e os periódicos científicos passam a exigir a elaboração do PGD. Essa dificuldade pode ocorrer

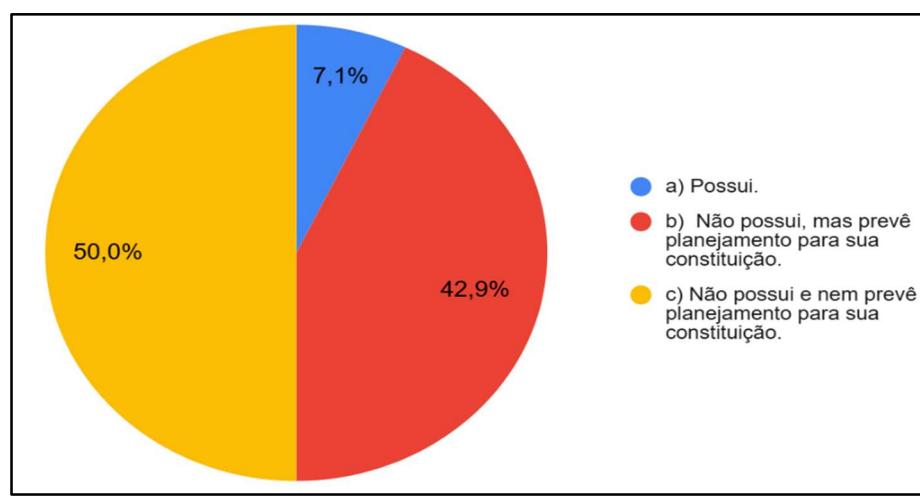
principalmente no contexto internacional, no qual a exigência da elaboração do PGD já é uma realidade.

O PGD é um documento que pode gerar benefícios em torno da preservação dos dados, visibilidade e reproduzibilidade da pesquisa, além de criar diretrizes para o arquivamento, uso e segurança dos dados (Smale *et al.*, 2018; Príncipe *et al.*, 2020). Nesse contexto, Silva (2016) relata que alguns pesquisadores ainda têm dificuldades para proceder com o preenchimento do PGD e enfatiza a necessidade da criação desse serviço pelos profissionais bibliotecários para auxiliar os pesquisadores em seu preenchimento.

Ao serem questionados se **o sistema de bibliotecas possui serviço de orientação para anonimização dos dados pessoais e sensíveis dos conjuntos de dados pertencentes as pesquisas desenvolvidas na universidade** (nona pergunta do questionário), apenas o SIBI11 informou possuir o serviço e complementa sua resposta afirmando que o serviço oferecido tem como objetivo o “Controle dos dados sensíveis através da análise/cadastro dos metadados e liberação pública das informações”.

Em contrapartida, seis (42,9%) respondentes afirmaram que estão prevendo planejamento para a concepção do serviço e sete (50,0%) não o possuem e nem preveem planejamento para o referido serviço. Os resultados demonstram que quase a totalidade dos sistemas de bibliotecas ainda não realiza o serviço, conforme o Gráfico 8.

Gráfico 8 - Orientação para anonimização dos dados

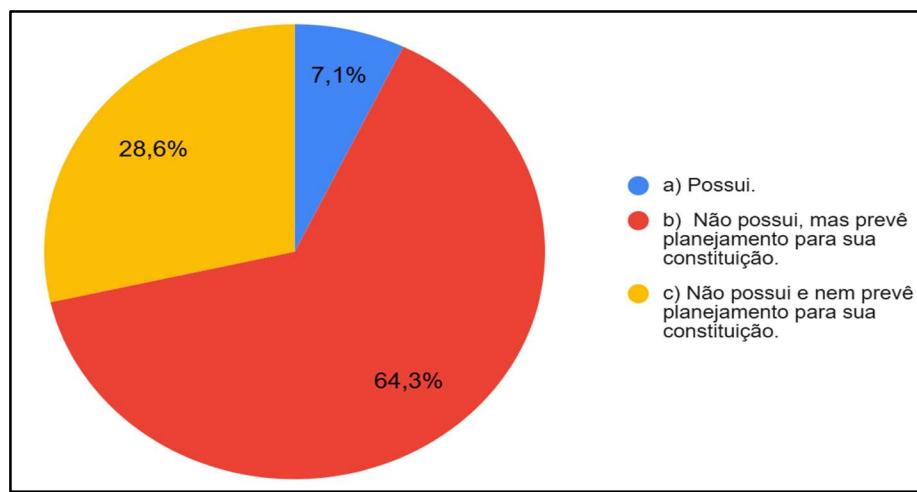


Serviços de anonimização dos dados pessoais e sensíveis são imprescindíveis nas pesquisas científicas, pois orientam os pesquisadores sobre quais dados poderão ser compartilhados com o público sem transgredir leis de proteção de dados vigentes em cada país.

Os sistemas de bibliotecas do Nordeste brasileiro que planejam disponibilizar o serviço, devem ter profissionais com conhecimento aprofundado sobre o principal documento que rege o tratamento e compartilhamento de dado pessoais e sensíveis no Brasil, neste caso a LGPD. Parcerias com outras instâncias da universidade como a área jurídica, podem ser uma boa estratégia para promover um serviço mais robusto para a comunidade científica.

A respeito do questionamento se o sistema de bibliotecas possui orientação para os pesquisadores da universidade encontrarem dados úteis para o desenvolvimento de suas pesquisas, apenas o SIBI1 respondeu possuir o serviço, descrevendo-o como “Treinamento em base de dados (pra turma ou grupo) e assessoria de pesquisa (orientação individual)”. Nove (64,3%) sistemas de biblioteca informaram não possuírem, mas preveem planejamento para a sua constituição, enquanto quatro (28,6%) respondentes afirmaram “não possui e nem prevê planejamento para sua constituição”, conforme o Gráfico 9.

Gráfico 9 - Orientação para encontrar dados úteis



Fonte: elaboração própria.

Apesar de somente um sistema de bibliotecas possuir o referido serviço, os outros dados da análise demonstram que uma parcela significativa dos sistemas de

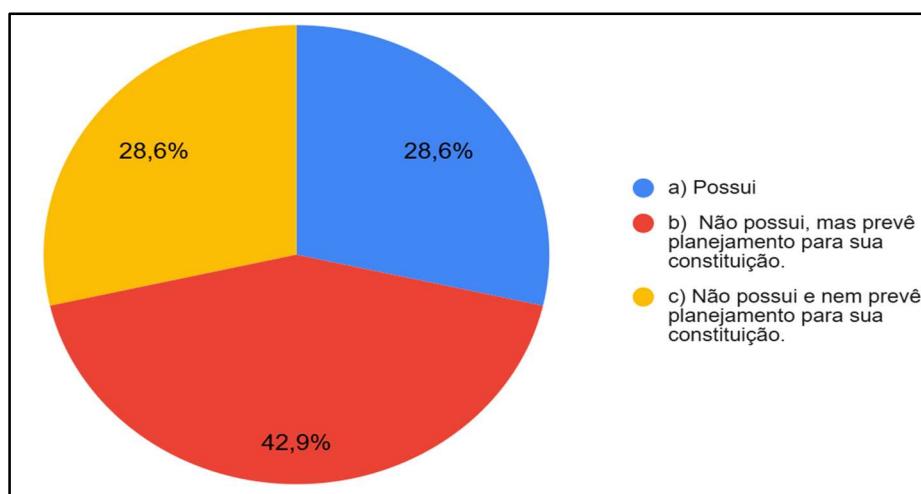
bibliotecas prevê o planejamento para a sua constituição, revelando, assim, um posicionamento positivo por parte desses sistemas de bibliotecas.

Serviços de orientação para que os pesquisadores a encontrarem dados úteis são excelentes estratégias para poupar o tempo do pesquisador e promover celeridade em suas pesquisas. No estudo desenvolvido por Pralle *et al.*, (2013) com a participação de setenta e três bibliotecas dos Estados Unidos, 93% dos entrevistados responderam possuir serviço para orientar os pesquisadores a localizarem e utilizarem dados, demonstrando uma grande adesão desse tipo de serviço nas bibliotecas que foram abordadas nesse estudo.

Antes de proceder com a análise do Gráfico 10, é importante ressaltar que a pergunta do item onze do questionário foi propositalmente construída em uma estrutura que engloba toda a instituição e não somente o sistema de bibliotecas, pois comprehende-se que serviços de repositórios de dados científicos são construções complexas, que exigem a participação de diversos componentes da instituição para a sua concepção.

Dando continuidade à análise, quando questionados se **a universidade possui repositório de dados científicos**, quatro (28,6%) sistemas de bibliotecas responderam que possuem o referido serviço, seis (42,9%) respondentes informaram não possuírem, mas preveem planejamento para sua constituição e quatro (28,6%) não possuem e nem preveem planejamento para a constituição”, conforme o Gráfico 10.

Gráfico 10 - Repositório de dados científico



Fonte: elaboração própria.

Os sistemas de bibliotecas que optaram pela opção “possui” descreveram a função do repositório de dados científicos como sendo:

Disponibilizar e divulgar dados e pesquisas desenvolvidos no âmbito da universidade (SIBI4).

O repositório de dados da instituição está aguardando a liberação de espaço de armazenamento a fim de que possa liberar o Dataverse para uso e depósito dos pesquisadores. Estamos em fase de implantação, por isso, apesar de o repositório de dados existir e já ter instalação pronta, estamos dependendo da liberação de mais espaço de armazenamento para iniciar o funcionamento da forma adequada. O repositório de dados irá abrigar uma diversificada tipologia, tais como números, modelos de questionários, imagens, avaliações, transcrições de entrevistas, textos, vídeos, áudio, protocolos, banco de dados, planilhas, algoritmos, animações, sequências genéticas, modelos, metodologias, simulações, poligonais, coordenadas geográficas etc. (SIBI8).

Armazenar, divulgar e democratizar a informação científica (SIBI11).

O Repositório tem o propósito de reunir, armazenar, organizar, preservar e disseminar a produção científica e intelectual da comunidade universitária (docentes, pesquisadores, técnicos e alunos de graduação e pós-graduação stricto e lato sensu) pertencente à Universidade [...] (SIBI14).

Embora o SIBI4 tenha afirmado possuir repositório de dados científicos, ao realizar pesquisa no *site* da instituição para a comprovação do serviço, apenas foi identificado o Repositório Institucional de Publicações Científicas (Atena)¹⁴⁴. Cenário semelhante também foi constatado na pesquisa realizada no *site* do SIBI11 e do SIBI14, onde foram identificados apenas os repositórios institucionais de publicação científica¹⁴⁵. Quanto ao SIBI8, na pesquisa realizada no *site* da instituição foi possível identificar a existência do Repositório de Dados Científico¹⁴⁶, exposto na Figura 37.

¹⁴⁴ Ver informações em: <https://repositorio.ufpe.br/>

¹⁴⁵ Ver informações em: <https://repositorio.ufersa.edu.br/home> e <https://repositorio.unilab.edu.br/jspui/>

¹⁴⁶ Ver informações em: <https://dadosdepesquisa.ufc.br/dataverse/redam>

Figura 37 - Repositório de Dados de Pesquisa Antônio Martins Filho (REDAM)



Fonte: *site* do Repositório de dados do SIBI4.

O REDAM foi desenvolvido com *software* de código-fonte aberto DataVERSE e tem como objetivo a promoção do compartilhamento, reuso, citação, visibilidade, preservação e segurança dos dados científicos da comunidade da instituição. O gerenciamento do REDAM fica sob a liderança do Sistema de Bibliotecas com a colaboração de outras instâncias da universidade (UFC, 2025).

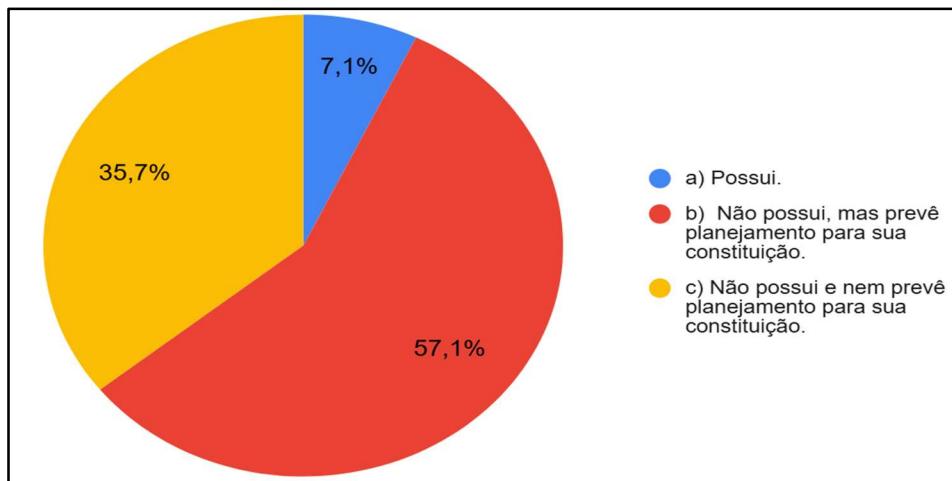
Ao término da análise, constatou-se que apenas um (7,1%) sistema de bibliotecas possui o serviço e que sete (50,0%) não possuem o serviço e nem preveem planejamento para sua formulação. Apesar do questionário conter a conceituação sobre repositório institucional de publicações científicas e repositório institucional de dados científicos, percebe-se que existem coordenações e diretorias dos sistemas de bibliotecas que se equivocaram quanto à oferta desse serviço em seus sistemas de bibliotecas.

Nesse sentido, é importante ressaltar que a gestão de repositórios de dados científicos é constituída por diversos elementos que se complementam para o funcionamento dessa infraestrutura eletrônica, entre alguns desses elementos podemos citar: aspectos jurídicos (LGPD e Lei de Acesso à Informação), políticas do repositório, licenças específicas para uso dos dados (direitos autorais), curadoria e ciclo de vida dos dados, certificações de repositório confiável, preservação de dados, manipulação da linguagem de programação do *software*, backup e proteção dos dados. Essa complexidade de elementos descritos anteriormente exige uma equipe multidisciplinar para obter êxito no funcionamento do repositório de dados científicos (Silva, 2016; Sayão; Sales, 2022).

Ao serem questionados se **O sistema de bibliotecas possui orientação para indicar repositórios de dados científicos para o depósito dos conjuntos de dados dos pesquisadores da universidade**, apenas um (7,1%) sistema de

bibliotecas afirmou possuir o serviço. Oito (57,1%) unidades não possuem, mas preveem planejamento para sua constituição, enquanto cinco (35,7%) unidades não possuem e nem preveem o planejamento para sua constituição”, conforme o Gráfico 11.

Gráfico 11 - Orientação para indicar repositórios de dados científicos



Fonte: elaboração própria.

O SIBI1, único sistema de bibliotecas a possuir o serviço, informou que “Essa orientação é ofertada durante o treinamento em base de dados e na assessoria de pesquisa”, revelando a existência do serviço e a regularidade do evento para estas orientações.

A indicação de repositórios de dados científicos requer que o sistema de bibliotecas estruture o serviço tomando como base algumas informações relevantes para a sua concepção. Nessa perspectiva, é importante conhecer as políticas de acesso e preservação do repositório; quem é a instituição ou grupo responsável pelo gerenciamento do repositório, pois isso ajuda a compreender por quanto tempo o serviço ficará disponível; verificar se utiliza identificadores persistentes; quais os formatos de arquivos que são aceitos para o depósito; se possui protocolos de interoperabilidade e padrão de metadados que possa descrever o conjunto de dados de forma adequada (Silva, 2016; Sayão; Sales, 2022).

Outro ponto importante é entender as características de cada categoria de repositório e quais as suas funções. Por exemplo, repositórios temáticos são excelentes para promover a visibilidade dos dados em comunidades de disciplinas específicas ou em determinadas áreas do conhecimento, ao passo que os repositórios

multidisciplinares possuem uma capacidade maior de abrangência, porém para um público diverso (Silva, 2016).

Por fim, solicitou-se que os sistemas de bibliotecas descrevessem, caso possuíssem algum serviço voltado para o gerenciamento de dados científicos ou planejamento para esse propósito além dos previstos neste questionário. Entre os quatorze sistemas de bibliotecas, oito não responderam, duas instituições responderam que não possuíam outros serviços, e quatro apresentaram informações que estão descritas no Quadro 16.

Quadro 16 – Relatos dos sistemas de bibliotecas

Sistema de Bibliotecas	Relato
SIBI2	Enquanto não temos um repositório de dados específico da instituição, orientamos os pesquisadores que querem ou precisam publicizar seus dados de pesquisa a utilizarem alguns repositórios de uso coletivo
SIBI3	Os servidores do Repositório Institucional têm participado da RENERE onde a discussão sobre a implantação e importância dos Repositório de dados tem sido discutida. A partir disto pretendemos iniciar um diálogo com a Administração Superior para ampliar a discussão e construir perspectivas
SIBI8	Apesar de não ter criado oficialmente vários dos serviços mencionados neste questionário, a universidade atende, em certa medida, as demandas que chegam relacionadas aos dados de pesquisa. Essas demandas são direcionadas para a Biblioteca e para a Bibliotecária especializada nessa área e que é responsável pelo planejamento dos serviços que serão oficialmente divulgados tão logo o repositório de dados seja lançado na universidade. Diversas atividades já são executadas por essa profissional de forma extraoficial, então, apenas estamos trabalhando para que o repositório seja liberado em breve para a comunidade
SIBI9	No momento temos apenas a Biblioteca Digital de Teses e Dissertações (BDTD), estamos na construção do Repositório Institucional (RI)

Fonte: dados da pesquisa (2025).

O relato do SIBI2 contradiz a resposta apresentada na pergunta anterior, referente à orientação para indicação de repositórios de dados científicos para o depósito do conjunto de dados dos pesquisadores da universidade. Isso ocorre porque o SIBI2 relatou “não possuir, mas prevê planejamento para sua constituição”. De

acordo com o exposto não foi possível compreender se o referido Sistema de Bibliotecas possui ou não o serviço.

As informações fornecidas pelo SIBI3 demonstram que o Sistema de Bibliotecas através da Rede Nordeste de Repositórios está buscando as informações necessárias para posteriormente dialogar com instâncias superiores para a implantação do seu repositório de dados científicos. A Conduta do SIBI3 corrobora com as premissas de Jones, Pryor e White (2013) no tocante à necessidade de envolver outras instâncias da universidade, principalmente a alta gestão, para a concepção de SPGDC.

Através das informações expostas pelo SIBI8, comprehende-se que diversos serviços de gestão de dados científicos são realizados por uma bibliotecária que possui conhecimento na temática. Contudo, os serviços ainda não constam no catálogo de serviços do Sistema de Bibliotecas. Nesse sentido, é fundamental que esses serviços sejam oficializados e incluídos no catálogo de serviço do Sistema de Bibliotecas para o conhecimento da comunidade acadêmica.

Já o serviço relatado pelo SIBI9 não está condizente com um serviço de gerenciamento de dados científicos, pois Bibliotecas Digitais de Teses e Dissertações e Repositórios Institucionais de Publicações Científicas são plataformas destinadas ao gerenciamento de produções científicas no contexto do Acesso Aberto, assim a presente resposta revela um desconhecimento sobre a temática por parte do(a) respondente.

8 CONSIDERAÇÕES FINAIS

As transformações ocorridas no campo científico centrado em dados, nas últimas décadas, estão revolucionando a produção e a comunicação científica. Os dados científicos produzidos nas IES que outrora eram descartados ou armazenados de forma indevida, passaram a ser insumos valiosos para o avanço científico devido ao reconhecimento das suas aplicações na retroalimentação da ciência.

Nesse seguimento, as agências de fomento à pesquisa e os periódicos científicos passaram a exigir o compartilhamento dos conjuntos de dados das produções científicas tornando-os ainda mais significativos para a comunicação e a produção científica. Entretanto, essa decisão demanda mais atenção ao tratamento para a disponibilização desses dados científicos, pois este é um processo complexo que requer técnicas específicas, infraestruturas tecnológicas, recursos financeiros e pessoas capacitadas para o seu devido gerenciamento.

Consubstancialmente, esta nova configuração no campo científico, resultou na necessidade do desenvolvimento de serviços e produtos que auxiliem os pesquisadores no gerenciamento dos conjuntos de dados das suas produções científicas e foi nesse contexto multifacetado que buscou-se entender como os sistemas de bibliotecas das universidades federais do Nordeste brasileiro vêm se preparando para promover a gestão dos dados científicos das suas instituições. Com o objetivo de responder a esse questionamento, foi realizado um diagnóstico que detectou quais serviços e produtos são ofertados pelos sistemas de bibliotecas das universidades federais do Nordeste brasileiro.

Sob este prisma, os objetivos específicos da pesquisa permitiram alcançar os seguintes caminhos e conclusões:

Mapear a partir da literatura científica, quais os serviços e produtos para o gerenciamento de dados científicos: permitiu através do conhecimento científico identificar o que é gestão de dados científicos e quais os SPGDC existentes; as dificuldades encontradas pelas bibliotecas no planejamento e ofertas dos SPGDC; a necessidade das bibliotecas estabelecerem parcerias com outros setores da IES para a oferta desses serviços; necessidade de capacitação do profissional bibliotecário para atuar como bibliotecário de dados e a produção de informações para a criação do questionário empregado para alcançar o objetivo específico 3.

Averiguar nos sites de bibliotecas de instituições de ensino superior estrangeiras a existência de serviços e produtos destinados ao gerenciamento de dados científicos, contribuindo com informações complementares para elaboração do diagnóstico: possibilitou identificar nos sites das bibliotecas de *Harvard University, MIT e University of Cambridge*, quais são os SPGDC e como estes são ofertados; permitiu compreender que muitos dos serviços ofertados pelas bibliotecas são provenientes de parcerias estabelecidas com outros setores da instituição e externos à instituição; constatou-se a existência de equipes com conhecimentos específicos em gestão de dados para trabalhar com o planejamento e a execução dos SPGDC nas bibliotecas e subsidiou também a produção do questionário que atenderia ao objetivo específico 3.

Identificar em qual estágio se encontram os produtos e serviços para gestão de dados científicos nos sistemas de bibliotecas das universidades federais do Nordeste brasileiro: revelou, por meio da aplicação do questionário da pesquisa quais SPGDC são ofertados por esses sistemas de bibliotecas e quais SPGDC estão em planejamento.

Os resultados da presente pesquisa revelaram uma baixa oferta de SPGDC nos sistemas de bibliotecas das universidades federais do Nordeste brasileiro. Sobre os onze SPGDC perguntados no questionário, seis não são ofertados por nenhum sistema de bibliotecas, quatro SPGDC são ofertados, cada um, por apenas um sistema de bibliotecas e um SPGDC é ofertado por três sistemas de bibliotecas simultaneamente. Nessa perspectiva, os dados revelados das bibliotecas estrangeiras, em termos comparativos, demonstram que os sistemas de bibliotecas das universidades federais do Nordeste brasileiro necessitam promover uma mudança de postura em relação a oferta de SPGDC. Os dados descritos podem ser visualizados no Quadro 17.

Quadro 17 – Bibliotecas Estrangeiras X Bibliotecas do Nordeste brasileiro

SPGDC	Bibliotecas Estrangeiras	Bibliotecas Universitárias do Nordeste brasileiro
Política para GDC	Duas possuem	Não possui
Divulgação informações sobre GDC	Todas possuem	Não possui
Palestras sobre GDC	Todas possuem	Três bibliotecas possuem

Capacitação de bibliotecários sobre GDC	Todas possuem	Não possui
Princípios FAIR	Todas possuem	Não possui
Ciclo de vida dos dados	Todas possuem	Não possui
Plano de Gestão de Dados (PGD)	Todas possuem	Não possui
Anonimização de dados pessoais e sensíveis	Todas possuem	Uma biblioteca possui
Encontrar dados úteis	Todas possuem	Uma biblioteca possui
Repositório de dados científicos	Todas possuem	Uma biblioteca possui
Indicar repositório de dados científicos	Todas possuem	Uma biblioteca possui

Fonte: dados da pesquisa (2025).

A quantidade de sistema de bibliotecas que não possui SPGDC e nem prevê planejamento para a sua formulação também alcançou números elevados após a finalização da análise das respostas: 57,1% na pergunta dois, 57,1% na pergunta três, 42,9% na pergunta quatro, 64,9% na pergunta cinco, 57,1% na pergunta seis, 64,3% na pergunta sete, 57,1% na pergunta oito, 50,0% na pergunta nove e 50,0% na pergunta onze. Os resultados demonstraram que a comunidade científica dessas IES além de não ter a disponibilização SPGDC, também ficará por um certo tempo sem o planejamento por parte dos seus sistemas de bibliotecas para conceber SPGDC.

Outro dado revelado pela pesquisa foi o desconhecimento de algumas coordenações e diretorias dos sistemas de bibliotecas sobre o que pode ser considerado um SPGDC. O fato ficou constatado mediante algumas respostas presentes no questionário, onde foram cedidas informações divergentes dos SPGDC perguntados, como o envio do *link* da política do repositório institucional de publicações científicas e o *link* da página do repositório institucional de publicações científica.

Os resultados desta pesquisa permitiram inferir que os sistemas de bibliotecas das universidades federais do Nordeste brasileiro investigados carecem de efetivar as ofertas de SPGDC e os sistemas que até a realização da pesquisa não têm como pauta o planejamento de SPGDC precisam incluir-se nas discussões sobre a temática.

Para promover a modificação desse cenário conforme o que consta na literatura científica (teoria) e o que foi investigado nos *sites* de bibliotecas estrangeiras (prática) nesta pesquisa, faz-se necessário o estabelecimento de capacitações em

GDC para os bibliotecários, criação de políticas locais com diretrizes para GDC, conhecer quais SPGDC são ofertados por outros sistemas de bibliotecas, aplicar questionários para conhecer as necessidades de sua comunidade em relação a oferta de SPGDC, realizar eventos para promover informações e discussões acerca da Ciência Aberta com foco no gerenciamento de dados científicos, estabelecer parcerias com diferentes instâncias da instituição e inclusive com outras instituições, disponibilização de verbas direcionadas para SPGDC, identificar os requisitos dos periódicos científicos e agências de fomento à pesquisa sobre o compartilhamento de dados e formulação do PGD e sensibilizar a alta gestão sobre a importância da oferta de SPGDC na instituição.

Diante desse complexo cenário, o percurso para estabelecer as diretrizes citadas anteriormente exigirá dos sistemas de bibliotecas das universidades federais do Nordeste brasileiro análises e reformulações de suas práticas existentes. Essa mudança de postura em relação ao gerenciamento dos dados científicos em sua coleta, tratamento, arquivamento, compartilhamento e preservação, poderá proporcionar para as universidades, pesquisadores, produções científicas e a ciência diversos benefícios, como:

- Transparência no processo científico: possibilidade de conhecer quais etapas foram realizadas pelo pesquisador no decurso do seu projeto científico;
- Reprodutibilidade da pesquisa: acesso aos conjuntos de dados para reprodução da pesquisa científica com o objetivo de promover o rigor científico;
- Credibilidade e integridade científica: proporciona a identificação dos processos pelos quais o pesquisador chegou até os resultados da sua pesquisa, viabilizando a detectação de fraudes, erros metodológicos e fabricações de dados para a manipulação de resultados;
- Reuso de dados: permite a reutilização de dados de estudos já realizados para pular etapas no processo de construção de outras pesquisas;
- Economia de recursos públicos: evitar que agências de fomento à pesquisa financiem projetos científicos para coletar dados que já foram coletados;

- Visibilidade científica: pesquisadores que tornam os seus conjuntos de dados acessíveis conferem maior número de citações para as suas produções científicas;
- Avanço científico: geração de novas pesquisas a partir de dados disponibilizados, como por exemplo os dados de estudos compartilhados sobre a covid-19, possibilitaram a criação de vacinas e medidas de combate ao vírus no período pandêmico iniciado em 2020 com mais rapidez;
- Tornar os dados em dados *FAIR*: proporciona mecanismos para que os dados se tornem Encontráveis, Acessíveis, Interoperáveis e Reutilizáveis. Nesse sentido, os dados passam a ser açãoáveis por máquina, permitindo o uso de metadados específicos para recuperação dos dados e possibilitando a comunicação com plataformas distintas para o compartilhamento de dados;
- Atender às exigências dos periódicos científicos e agências de fomento à pesquisa acerca da GDC: tratamento adequado dos dados científicos em seu ciclo de vida e construção de PGD;
- Preservação digital: proporciona a preservação dos dados independentemente da obsolescência da infraestrutura tecnológica, assim permitindo o acesso aos dados a longo prazo e o seu reuso.

Diante dos fatos supracitados, comprehende-se que as bibliotecas universitárias possuem responsabilidade e capacidade para liderar e desempenhar o papel de suporte no gerenciamento de dados científicos no ambiente das IES, pois esses órgãos estão intrinsecamente ligados ao suporte da pesquisa científica nas universidades e já possuem experiência com o tratamento das produções científicas tradicionais (livros, artigos de periódicos, artigos de evento, teses e dissertações). Apesar do diagnóstico realizado na presente pesquisa revelar baixa oferta de SPGDC pelos sistemas de bibliotecas das universidades do Nordeste brasileiro, as informações identificadas na literatura científica e nos sites das bibliotecas de *Harvard University*, *MIT* e *University of Cambridge* demonstraram viabilidade para oferta de SPGDC, principalmente com a capacitação dos profissionais bibliotecários por meio da disponibilização de recursos financeiros para iniciar esse processo.

Tenopir *et al.* (2016) destacam uma correlação direta entre a capacitação dos profissionais bibliotecários e a oferta de serviços de gerenciamento de dados

científicos. Sob essa ótica, a baixa oferta de SPGDC nas bibliotecas do Objetivo específico 3 pode estar diretamente correlacionada com a ausência de suporte financeiro para capacitar os bibliotecários em GDC. Pois, ao serem questionados, registrou-se consenso entre os respondentes, ao afirmarem que não possuem este serviço. Contudo, esclarece-se que a pesquisa não se objetivou identificar os motivos pelos quais os SPGDC não são ofertados, mas recomenda-se que outros estudos sejam desenvolvidos com o objetivo de identificar essa correlação e suas implicações.

Em relação às limitações da pesquisa, considera-se que o estudo poderia apresentar informações adicionais caso estabelecesse uma investigação complementar nos *sites* dos sistemas de bibliotecas das universidades federais do Nordeste brasileiro, com o objetivo de identificar serviços e produtos que não estavam presentes no questionário, assim como a averiguação da existência de Planos de Dados Abertos e a sua contribuição para os sistemas de bibliotecas procederem com a gestão dos dados científicos da instituição. Contudo, devido ao prazo de entrega da dissertação, o processo foi considerado inviável.

REFERÊNCIAS

ALBAGLI, S. Ciência aberta em questão. *In: ALBAGLI, S.; MACIEL, M. L.; ABDO, A. H. Ciência aberta, questões abertas.* Brasília, DF: IBICT; Rio de Janeiro: UNIRIO, 2015. p. 9-26. Disponível em: https://livroaberto.ibict.br/bitstream/1/1060/1/Ciencia%20aberta_questoes%20aberta_s_PORTUGUES_DIGITAL%20%285%29.pdf. Acesso em: 5 ago. 2020.

ALBAGLI, S.; CLINIO, A.; RAYCHTOCK, S. Ciência aberta: correntes interpretativas e tipos de ação. *Liinc em Revista*, Rio de Janeiro, v. 10, n. 2, nov. 2014. Disponível em: <http://revista.ibict.br/liinc/index.php/liinc/article/view/749/512>. Acesso em: 11 ago. 2020.

ALMEIDA, A. C. L.; SANCHEZ-TARRAGO, N. Comunidade de prática: diagnóstico em redes regionais de repositórios no Brasil. *In: ENCONTRO NACIONAL DE PESQUISA E PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO*, 2023. Anais [...] XXIII Encontro Nacional de Pesquisa e Pós-graduação em Ciência da Informação, 2023.

AMORIM, R. C. *et al.* A comparison of research data management platforms: architecture, flexible metadata and interoperability. *Universal Access. The information Society*, New York, v. 16, n. 4, p. 851-862, 2017. Disponível em: <https://repositorioaberto.up.pt/bitstream/10216/111537/2/229906.pdf> Acesso em 10 ago. 2020.

ANCION, Z. *et al.* **Action plan for diamond open access.** [s. l.]: Scie Europe, 2022. Disponível em: <https://www.scienceeurope.org/our-resources/action-plan-for-diamond-open-access/>. Acesso em: 10 dez. 2024.

ARXIV. [S. l.], [201-]. Disponível em: <https://arxiv.org/>. Acesso em: 8 jul. 2021.

ATKINS, D. E. *et al.* **Revolutionizing Science and Engineering through Cyberinfrastructure. Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure.** Arlingto: Directorate for Computer and Information Science and Engineering, National Science Foundation, 2003. Disponível em: https://www.researchgate.net/publication/224361906_Revolutionizing_Science_and_Engineering_Through_Cyberinfrastructure. Acesso em: 20 out. 2020.

AUSTRALIAN RESEARCH DATA COMMONS (Austrália). **Research data management framework for institutions.** [Sydney]: National Research Infrastructure for Australia, 2023. Disponível em: <https://zenodo.org/records/8433246>. Acesso em: 11 fev. 2025.

BALL, A. **Review of Data Management Lifecycle Models.** Bath: University of Bath, 2012. Disponível em: <https://purehost.bath.ac.uk/ws/portalfiles/portal/206543/redm1rep120110ab10.pdf>. Acesso em: 5 jul. 2024.

BANDEIRA, P. M. **Movimento de Acesso Aberto no Brasil**: contribuição do Instituto Brasileiro de Informação em Ciência e Tecnologia a partir da implementação do Sistema Eletrônico de Editoração de Revistas. 2017. Dissertação (Mestrado em Ciência da Informação) – Centro de Ciências Sociais, Universidade Federal da Paraíba, João Pessoa, 2017. Disponível em: <https://repositorio.ufpb.br/jspui/bitstream/te/9702/2/ArquivoTotal.pdf>. Acesso em: 15 ago. 2021.

BARBOSA W. L.; LYRA R. S. **Governança de dados Módulo 2: princípios, importância e desafios do gerenciamento de dados**. Brasília, DF: ENAP, 2019. Disponível em: <https://repositorio.enap.gov.br/bitstream/1/5008>. Acesso em 20 mar. 2025.

BARRAVIEIRA, B. *et al.* O Alinhamento das revistas brasileiras sobre Medicina Tropical e doenças infecciosas e parasitárias as práticas. In: SHINTAKU, M.; SALES, L. F.; COSTA, M. (org.). **Tópicos sobre dados abertos para editores científicos**. Botucatu: ABEC, 2020. p. 215-230.

BARRETO SEGUNDO, J. de D. *et al.* Relações entre Acesso Aberto, QUALIS CAPES e desempenho de citação (Índices h, e, AW e h Anual) em periódicos científicos brasileiros de Ciência da Informação – estudo documental exploratório. **Informação & Sociedade**: estudos, João Pessoa, v. 30, n. 1, 5 mar. 2020. Disponível em: <https://periodicos.ufpb.br/ojs2/index.php/ies/article/view/47362>. Acesso em: 20 maio 2020.

BARRETO, A. de A. A condição da informação. **Revista São Paulo em Perspectiva**, São Paulo, v. 16, n. 3, p. 67-74, jul. 2002. Disponível em: http://www.scileio.br/scileio.php?script=sci_arttext&pid=S010288392002000300010&ing=en&nrm=iss. Acesso em: 14 ago. 2020.

BERLIM (Alemanha). **Berlin declaration on open access to knowledge in the science and humanities**. Open Access MaxPlanck-Gesellschaft, 2003. [S. l.: s. n.], 2003. Disponível em: <https://openaccess.mpg.de/BerlinDeclaration>. Acesso em: 20 jul. 2020.

Bertin, P.R.B. *et al.* A Parceria para Governo Aberto como plataforma para o avanço da Ciência Aberta no Brasil. **Transinformação**, v.31, 2019. Disponível em: <https://www.scielo.br/j/tinf/a/hZGKbLBrvv3KjBFNnZ65qDN/>. Acesso em: 20 jul. 2024.

BORKO, H. Information science: what is it? **American Documentation**, v.19, n.1, p.3-5, Jan. 1968. Disponível em: <https://onlinelibrary.wiley.com/doi/epdf/10.1002/asi.5090190103>. Acesso em: 20 out. 2024.

BONETI, Lindomar Wessler. **Políticas Públicas por Dentro**. 2^a Edição. Editora Unijuí: Ijuí, 2010.

BORGMAN, C. L. The conundrum of sharing research data. **Journal of the Association for Information Science and Technology**, New York, v. 63, n. 6, p. 1059-1078, jun. 2012. Disponível em: <https://dl.acm.org/citation>.

cfm?id=2222887. Acesso em: 23 nov. 2017.

BOURDIEU, P. **Usos sociais da ciência**: por uma sociologia clínica do campo científico. São Paulo: Ed. UNESP, 2004.

BRASIL. Lei nº 12.527, de 18 de janeiro de 2012. Regula o acesso a informações previsto no inciso XXXIII do art. 5º no inciso II do § 3º do art. 37 e no § 2º do art. 216 da Constituição Federal; altera a Lei nº 8.112, de 11 de dezembro de 1990; revoga a Lei nº 11.111, de 5 de maio de 2005, e dispositivos da Lei nº 8.159, de 8 de janeiro de 1991; e dá outras providências. **Diário Oficial da União**: seção 1, Brasília, DF, ano 148, n. 221-A, p. 1-4, 18 nov. 2011. Disponível em: <https://pesquisa.in.gov.br/imprensa/jsp/visualiza/index.jsp?data=18/11/2011&jornal=1000&pagina=1&totalArquivos=12>. Acesso em: 10 abr. 2021.

BRASIL. Decreto nº. 8.777, de 11 de maio de 2016. Institui a Política de Dados Abertos do Poder Executivo Federal. **Diário Oficial da União**, Brasília, 11 maio 2016. Disponível em: http://www.planalto.gov.br/CCIVIL_03/_Ato2015-2018/2016/Decreto/D8777.htm. Acesso em: 10 abr. 2021.

BRASIL. Lei nº 13.709, de 14 de agosto de 2018. Dispõe sobre a proteção de dados pessoais e altera a Lei nº 12.965, de 23 de abril de 2014 (Marco Civil da Internet). **Diário Oficial da União**, Brasília, DF, 14 ago. 2018. Disponível em: https://www.in.gov.br/materia-/asset_publisher/Kujrw0TzC2Mb/content/id/36849373/do1-2018-08-15-lei-no-13-709-de-14-de-agosto-de-2018-36849337. Acesso em 10 abr. 2021.

BRASIL. Lei nº 9.610, de 19 de fevereiro de 1998. Altera, atualiza e consolida a legislação sobre direitos autorais e dá outras providências. **Diário Oficial [da] República Federativa do Brasil**, Brasília, DF, 20 fev. 1998. Disponível em: https://www.planalto.gov.br/ccivil_03/leis/l9610.htm. Acesso em: 20 jul. 2021.

BRASIL. Ministério da Educação. **E-MEC**. Versão origin-master-778. Brasília, DF: MEC, 2024.

BRISBANE. **Brisbane Declaration**. BRISBANE: [s. n.], 2008.

BUCKLAND, M. K. Information as thing. **Journal of the American Society for Information Science**, New York, v. 45, n.5, p. 351-360, 1991. Disponível em: <https://www.cin.ufpe.br/~cjgf/TECNOLOGIA%20-%20material%20NAO-CLASSIFICADO/BUCKLAND%20Information%20as%20thing>. Acesso em: 18 out. 2021.

BUDAPEST open access initiative. **BOAI**, Hungria, 2002. Disponível em: <http://www.budapestopenaccessinitiative.org/read>. Acesso em 20 jul. 2020.

CAMPÊLO, L. R. R. R.; BARRETO NETO, V. C. Comparando Softwares gratuitos para criação de repositórios de dados abertos. **Ciência da Informação**, Brasília, DF, v. 48, n. 3, 20 mar. 2020. Disponível em: <http://revista.ibict.br/ciinf/article/view/5004/4467>. Acesso em: 9 ago. 2020.

CANADA. Government of Canada. **2018-2020 National Action Plan on Open Government**. Ottawa: [s. n.], 2018. Disponível em: <https://open.canada.ca/en/content/canadas-2018-2020-national-actionplan-open-government>. Acesso em: 6 maio 2021.

CAPURRO, R. et al. O conceito de informação. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 12, n. 1, p. 148-207, nov. 2007. Disponível em: <https://periodicos.ufmg.br/index.php/pci/article/view/22360/17954>. Acesso em: 2 ago. 2020.

CARDOSO, G.; JACOBETTY, P.; DUARTE, A. **Para uma Ciência Aberta**. Lisboa: Mundos Sociais, 2012. Disponível em: https://www.researchgate.net/publication/235994691_Para_Uma_Ciencia_Aberta. Acesso em: 4 abr. 2021.

CASTELLS, M. **A sociedade em rede**. São Paulo: Paz e Terra, 1999.

CENTRE FOR OPEN SCIENCE (COS). **Strategy for Culture Change**. 2024. Disponível em: <https://www.cos.io/about?hsCtaTracking=8b9438d6-1d53-4a72-9376-e23d2e5e5209%7C515640de-23b0-4a10-86ef-7d039af698f3>. Acesso em: 21 dez. 2023.

CENTRO DE ESTUDOS AVANÇADOS EM ECONOMIA APLICADA; CONFEDERAÇÃO DA AGRICULTURA E PECUÁRIA DO BRASIL. **PIB do agronegócio brasileiro**. São Paulo: [CEPEA], 2024. Disponível em: https://cepea.esalq.usp.br/upload/kceditor/files/CT-PIB-AGRO_26.MAR.24.pdf. Acesso em: 20 abr. 2024.

CÉSAR JÚNIOR, R. M. Apresentação à edição brasileira. In: HEY, T.; TANSLEY, S.; TOLLE, K. (org.). **O quarto paradigma**: cfm?id=2222887. Acesso em: 23 nov. 2017.

CIUFFO, L. N.; FELICISSIMO, C. H. A contribuição da RNP para a Ciência Aberta no Brasil. In: SILVA, F. C. C. da; STUEBER, K.; CARVALHO-SEGUNDO, W. L. R. de (org.). **Ciência Aberta no Brasil: conquistas e desafios**. Porto Alegre: Editora Letra1; São Paulo: SciELO, 2025. p. 181-200. DOI: <https://doi.org/10.5281/zenodo.15151031>. Disponível em: <https://zenodo.org/records/15183537>. Acesso em: 20 abr. 2025.

CONSELHO NACIONAL DE DESENVOLVIMENTO CIENTÍFICO E TECNOLÓGICO (CNPq). CNPq lançou consórcio CoNCienciA para incentivar a prática da Ciência Aberta. 2022. Disponível em: <https://www.gov.br/cnpq/pt-br/assuntos/noticias/cnpq-em-acao/cnpq-lancou-consorcio-conciencia-para-incentivar-a-pratica-da-ciencia-aberta-1> Acesso em: 12 jan. 2025.

CONSELHO NACIONAL DE DESENVOLVIMENTO CIENTÍFICO E TECNOLÓGICO (CNPQ). Histórico. 2020. Disponível em: <https://www.gov.br/cnpq/pt-br/acesso-a-informacao/institucional/historico> Acesso em: 12 jan. 2025.

CONSELHO NACIONAL DE DESENVOLVIMENTO CIENTÍFICO E TECNOLÓGICO (CNPq). Lattes Data. 2025. Disponível em: <https://lattesdata.cnpq.br/> Acesso em: 12 jan. 2025.

CONGRESSO MUNDIAL SOBRE RECURSOS EDUCACIONAIS ABERTOS (REA) DE 2012 UNESCO, PARIS, 20 A 22 DE JUNHO DE 2012. **Declaração REA de Paris 2012**. Disponível: https://unesdoc.unesco.org/ark:/48223/pf0000246687_por. Acesso em: 24 jul. 2021.

CONTROLADORIA GERAL DA UNIÃO. **Painel Monitoramento de Dados Abertos**. 2024. Disponível em: <https://centralpaineis.cgu.gov.br/visualizar/dadosabertos>. Acesso em: 20 dez. 2024.

CORE TRUST SEAL. **Trustworthy Digital Repositories Requirements 2023-2025: V01.00**. [s.l.]: Core Trust Seal, 2022. DOI: 10.5281/zenodo.7051011. Disponível em: <https://zenodo.org/records/7051012>. Acesso em: 20 jan. 2025.

COSTA, L. M. da; ARAUJO, A. dos S.; PAIXÃO, P. B. S. Análise das redes sociais do Sistema Integrado de Bibliotecas da Universidade Federal de Sergipe na perspectiva do marketing digital. **Ciência da Informação em Revista**, [s. l.], v. 11, p. e15847, 2024. DOI: 10.28998/cirev.2024v11e15847. Disponível em: <https://www.seer.ufal.br/index.php/cir/article/view/15847>. Acesso em: 15 fev. 2025.

COSTA, M. M. **Diretrizes para uma política de gestão de dados científicos no Brasil**. 2017. 288 f., il. Tese (Doutorado em Ciência da Informação) — Universidade de Brasília, Brasília, 2017. Disponível em: <https://repositorio.unb.br/handle/10482/24895>. Acesso em: 15 dez. 2024.

COSTA, Michelli. **Políticas promotoras de Ciência Aberta**. Brasília, DF. 13 nov. 2019. Apresentação em Slide Share. 19 slides. Color. 2º Encontro Capes de Ciência Aberta: direitos de propriedade intelectual e políticas institucionais. Disponível em: chrome-extension://efaidnbmnnibpcajpcglclefindmkaj/https://www.gov.br/capes/pt-br/centrais-deconteudo/20191113_2EncontroCAPESdeCienciaAberta_MichelleCosta.pdf. Acesso em: 15 dez. 2024.

COX, A. M.; PINFIELD, S. Research data management and libraries: current activities and future priorities. **Journal of Librarianship and Information Science**, v. 46, n. 4, p. 299–316, 28 2014. Disponível em: <http://lis.sagepub.com/cgi/doi/10.1177/0961000613492542>. Acesso em: 30 nov. 2023.

CREATIVE COMMONS. **About Licenses CC**. 21 nov. 2024. Disponível em: <https://creativecommons.org/share-your-work/cclicenses/>. Acesso em: 21 nov. 2024.

CUNHA, M. B. da. A biblioteca universitária na encruzilhada. **DataGramZero**: Revista de Ciência da Informação, Rio de Janeiro, v. 11, n. 6, dez. 2010. Disponível em: <http://icts.unb.br/jspui/handle/10482/14869>. Acesso em: 20 dez. 2023.

DANTAS, M. As rendas informacionais e a apropriação capitalista do trabalho científico e artístico. In: MARQUES, R. M. et al. (ed.). **A informação e o conhecimento sob as lentes do marxismo**. Rio de Janeiro: Garamond, 2014.

DECLARATION on Free Access to Law. **WorldLII**, [s. l.], [201-]. Disponível em: descobertas científicas na era da e-science. São Paulo: Oficina de Textos, 2011. p. 7-8. Disponível em: http://www.worldlii.org/worldlii/declaration/montreal_en.html. Acesso em: 15 jun. 2021.

DIAS, C. Políticas editoriais de compartilhamento de dados em periódicos brasileiros de ciências sociais aplicadas na coleção SciELO. **Encontros Bibli**: revista eletrônica de biblioteconomia e ciência da informação, Florianópolis, v. 29, p. 1-25, 2023. Disponível em: <https://periodicos.ufsc.br/index.php/eb/article/view/95038>. Acesso em: 22 jun. 2024.

DUDZIAK, E. Dados de pesquisa agora devem ser armazenados. **ABCD USP**, São Paulo, 2016. Disponível em: <https://www.aguia.usp.br/noticias/dados-materiais-metodos-revistas-exigem-dados-pesquisa-estejam-disponiveis/> Acesso em: 9 ago. 2020.

EDITAL Finep/Pcal/Xbdb 003/2009 foi prorrogado. **Blog do Kuramoto**, [s. l.], 2010. Disponível em: <https://kuramoto.wordpress.com/2010/04/27/edital-fineppcalxbdb-0032009-foi-prorrogado/>. Acesso em: 20 set. 2021.

FEARON JR, D. et al. **SPEC Kit 334: research data management services**. Washington, DC: ARL, 2013. Disponível em: <https://publications.arl.org/Research-Data-Management-Services-SPEC-Kit-334/>. Acesso em: 21 dez. 2024.

FECHER, B; FRIESIKE, S. Open Science: one term, five schools of thought. In: BARTLING, S; FRIESIKE, S. (Org.) **Opening science**. New York: Springer, 2014. p.17-47. Disponível em: <https://link.springer.com/book/10.1007/978-3-319-00026-8>. Acesso em: 20 jan. 2024.

FERGUSON, A. et al. Big data from small data: data-sharing in the 'long-tail' of neuroscience. **Nature Neuroscience**, New York, v. 17, n. 11, p. 1442-1447, nov. 2014. Disponível em: <https://www.nature.com/articles/nn.3838>. Acesso em: 23 out. 2023.

FERREIRA, V. B. **e-Science e políticas públicas para ciência, tecnologia e inovação no Brasil**: colaboração, infraestrutura e repercussão nos Institutos Nacionais de Ciência e Tecnologia da área de Nanotecnologia. 2016. Tese (Doutorado em Ciência da Informação) – Instituto de Ciência da Informação, Universidade Federal da Bahia, Salvador, 2016. Disponível em: <https://repositorio.ufba.br/ri/handle/ri/21149>. Acesso em: 20 jul. 2020.

FORNASIER, M. de O.; BECK, C. Cambridge Analytica: escândalo, legado e possíveis futuros para a democracia. **Revista Direito em Debate**, Unijuí, v. 29, n. 56, p. 182-195, jan./jun. 2020. Disponível em:

<https://www.revistas.unijui.edu.br/index.php/revistadireitoemdebate/article/view/1003>
 3. Acesso em: 27 fev. 2024.

FOSTER. **Open Science Definition**. 2014. Disponível em:
<https://web.archive.org/web/20190320013631/https://www.fosteropenscience.eu/foster-taxonomy/open-science-definition>. Acesso em: 21 set. 2021.

FOSTER. **OPEN Science Taxonomy**. [s. l.], [201-]. Disponível em:
https://www.fosteropenscience.eu/themes/fosterstrap/images/taxonomies/os_taxonomy.png. Acesso em: 21 jul. 2020.

FRANCE. Ministère de L'Enseignement supérieur de la Recherche et de L'Innovation. **National plan for open science**. Paris: Open Science, 2018. Disponível em: https://libereurope.eu/wp-content/uploads/2018/07/SO_A4_2018_05-EN_print.pdf. Acesso em: 6 maio 2021.

FREITAS-CAMPOS, F. *et al.* Contribuição do Instituto Brasileiro de Informação em Ciência e Tecnologia para a promoção do acesso aberto e da ciência aberta: análise de suas infraestruturas. *Revista Científica*, [Bogotá], v. 48, n. 3, p. 56-66, 2023. Disponível em:
<https://revistas.udistrital.edu.co/index.php/revcie/article/view/20924/19425>. Acesso em: 24 mar. 2024

FUNDAÇÃO DE CIÊNCIA, APLICAÇÕES E TECNOLOGIA ESPACIAIS. **Edital de chamada FINEP/PCAL/XBDB nº 002/2009**. [São Paulo]: FUNCATE, 2009.

FUNDAÇÃO OSWALDO CRUZ. Instituto de Comunicação e Informação Científica e Tecnológica em Saúde. Ferramenta online permite monitorar o avanço da Covid-19 no Brasil. **Portal Fiocruz Notícia**, Rio de janeiro, 7 abr. 2020. Disponível em: <https://portal.fiocruz.br/noticia/ferramenta-online-permite-monitorar-avanco-da-covid-19nobrasile#:~:text=Ferramenta%20online%20permite%20monitorar%20avan%C3%A7o%20da%C2%A0Covid%20no%20Brasil,-07%2F04%2F2020&text=Compartilhar%3A,no%20Brasil%20e%20no%20mundo>. Acesso em: 20 abr. 2024.

FUNDAÇÃO OSWALDO CRUZ. **A Fundação**. Rio de Janeiro, 11 jan. 2025. Disponível em: <https://portal.fiocruz.br/fundacao>. Acesso em: 11 jan. 2025.

FUNDAÇÃO OSWALDO CRUZ. **Política de Acesso Aberto ao Conhecimento da Fiocruz**. Rio de Janeiro, 31 mar. 2014. Disponível em:
https://portal.fiocruz.br/sites/portal.fiocruz.br/files/documentos/portaria_-_politica_de_acesso_aberto_ao_conhecimento_na_fiocruz.pdf Acesso em: 11 jan. 2025.

FUNDAÇÃO OSWALDO CRUZ. **Fiocruz lança Educare, novo espaço para educação aberta**. Rio de Janeiro, 24 set. 2019. Disponível em:
<https://campusvirtual.fiocruz.br/portal/?q=noticia/56759>. Acesso em: 11 jan. 2025.

FUNDAÇÃO OSWALDO CRUZ. **Ciência Aberta na Fiocruz**. Rio de Janeiro, 11 jan. 2025. Disponível em: <https://portal.fiocruz.br/educacao-aberta>. Acesso em: 11 jan. 2025.

FUNDAÇÃO OSWALDO CRUZ. Formulação modular em Ciência Aberta: dados abertos (aula 4 – Plano de Gestão de Dados). 2020. Disponível em: <https://mooc.campusvirtual.fiocruz.br/rea/ciencia-aberta/serie3/curso2/aula5.html>. Acesso em: 20 dez. 2021.

GIL, A. C. **Como elaborar projeto de pesquisa**. 5. ed. São Paulo: Atlas, 2010.

GO FAIR FUNDATION. **Implementation network**. 2024. Disponível em: <https://www.go-fair.org/implementation-networks/>. Acesso em: 21. Set. 2021.

GOLD, A. Cyberinfrastructure, data, and libraries, part 1: a cyberinfrastructure primer for librarians. **D-Lib Magazine**, [New York], v. 3, n. 9-10, p. 1-12, set./out. 2007. Disponível em: http://digitalcommons.calpoly.edu/cgi/viewcontent.cgi?article=1015&context=lib_dean. Acesso em: 15 maio 2021.

GOVERNMENT AUSTRALIAN. **Public Data Policy Statement**. 2015. Disponível em: https://www.finance.gov.au/sites/default/files/202210/aust_govt_public_data_policy_statement.pdf. Acesso em: 12 dez. 2024.

HARNAD, S. A Subversive Proposal. In: Okerson, A.; O'Donnell, J. (ed.). **Scholarly Journals at the Crossroads: A Subversive Proposal for Electronic Publishing**. Washington, DC: Association of Research Libraries, 1995. Disponível em: <https://eprints.soton.ac.uk/253351/>. Acesso em: 19 jan. 2024.

HARVARD UNIVERSITY. **Final NIH Policy for Data Management and Sharing**. 2023. Disponível em: <https://grants.nih.gov/grants/guide/notice-files/NOT-OD-21-013.html>. Acesso em: 20 dez. 2024.

HARVARD UNIVERSITY. **Research Data Resources**. 2025a. Disponível em: <https://osrds.library.harvard.edu/rd-resources/>. Acesso em: 15 fev. 2025.

HARVARD UNIVERSITY. **Research Data Service**. 2025b. Disponível em: <https://library.harvard.edu/services-tools/research-data-services>. Acesso em: 15 fev. 2025.

HARVARD UNIVERSITY. **Harvard Dataverse**. 2025c. Disponível em: <https://dataverse.harvard.edu/>. Acesso em: 15 fev. 2025.

HARVARD UNIVERSITY. **Gutman Library: Best Practices for Working with Data**. 2025d. Disponível em: <https://guides.library.harvard.edu/c.php?g=1312521&p=9748092> Acesso em: 20 maio. 2025.

HAYASHI, M. C. P. I.; GUIMARÃES, J. A. C. Dinâmicas sociais e princípios éticos: chaves de leitura sobre retratações em publicações científicas. **RDBCi: Revista Digital de Biblioteconomia e Ciência da Informação**, Campinas, SP, v. 22, n. 00, p. e024022, 2024. DOI: 10.20396/rdbc.i.v22i00.8676800. Disponível em: <https://periodicos.sbu.unicamp.br/ojs/index.php/rdbc/article/view/8676800>. Acesso em: 10 jan. 2025.

HEY, T.; TANSLEY, S.; TOLLE, K. Jim Gray on eScience: a transformed scientific method. In: HEY, T.; TANSLEY, S.; TOLLE, K. (ed.). **The fourth paradigm**: data intensive scientific discovery. Redmond: Microsoft Corporation, 2009. p. XVII-XXI. Disponível em: https://www.researchgate.net/publication/229529541_The_Fourth_Paradigm_Data-Intensive_Scientific_Discovery. Acesso em: 20 abr. 2020.

HICKS, D. *et al.* **Manifesto de Leiden sobre métricas de pesquisa**. [S. I.: s. n.], 2015. Disponível em: <http://www.leidenmanifesto.org/uploads/4/1/6/0/41603901/leiden-manifesto-portuguese-br-final.pdf>. Acesso em: 11 ago. 2020.

HIGGINS, S. Digital curation: the emergence of a new discipline. **The International Journal of Digital Curation**, [s. I.], v. 6, n. 2, p. 78-88, 2008. Disponível em: <http://www.ijdc.net/index.php/ijdc/article/view/184>. Acesso em: 22 jun. 2024.

HUANG, Y.; COX, A. M.; SBAFFI, L. Research data management policy and practice in Chinese University Libraries. **JASIST**, [s. I.], v. 72, n. 4, p. 493-506, 2021. Disponível em: <https://asistdl.onlinelibrary.wiley.com/doi/10.1002/asi.24413>. Acesso em: 20 dez. 2024.

IBICT. Dspace. **Portal do Gov.Br**: Ministério da Ciência, Tecnologia e Inovação, Brasília, DF, 26 jun. 2023a. Disponível em: <https://www.gov.br/ibict/pt-br/assuntos/tecnologias-para-a-informacao/dspace#:~:text=O%20DSpace%20%C3%A9%20um%20software,do%20software%20em%20%C3%A2mbito%20mundial>. Acesso em: 20 set. 2023.

IBICT. Histórico. **Portal do Gov.Br**: Ministério da Ciência, Tecnologia e Inovação, Brasília, DF, 15 out. 2018. Disponível em: <https://www.gov.br/ibict/pt-br/assuntos/tecnologias-para-a-informacao/ojs>. Acesso em: 20 set 2023.

IBICT. I Encontro sobre Gestão de Repositório Institucional. **IBICT**, Rio de Janeiro, 25 nov. 2009. Disponível em: <https://antigo.ibict.br/sala-de-imprensa/noticias/item/1280-i-encontro-sobre-gestao-de-repositorios-institucionais>. Acesso em: 20 set. 2021.

IBICT. **Manifesto de Acesso Aberto a Dados da Pesquisa Brasileira para Ciência Cidadã**. Rio de Janeiro: IBICT, 2016. Disponível em: <https://ibict.br/sala-de-imprensa/noticias/item/478-ibict-lanca-manifesto-de-acesso-aberto-a-dados-da-pesquisa-brasileira-para-ciencia-cidada>. Acesso em: 8 ago. 2020.

IBICT. OJS: Open Journal System. **Portal do Gov.Br**: Ministério da Ciência, Tecnologia e Inovação, Brasília, DF, 28 ago. 2023b. Disponível em:

<https://www.gov.br/ibict/pt-br/assuntos/tecnologias-para-a-informacao/ojs>. Acesso em: 20 set 2023.

IBICT. Confira como será a participação do IBICT na 15ª Conferência Lusófona de Ciência Aberta (ConfOA). Brasília, DF, 27 set. 2024. Disponível em: <https://www.gov.br/ibict/pt-br/central-de-conteudos/noticias/2024/setembro/confira-a-participacao-do-ibict-na-15a-conferencia-lusofona-de-ciencia-aberta-confoa>. Acesso em: 10 jan. 2025.

IBICT. Diadorim. Brasília, DF, 10 jan. 2025. Disponível em: <https://diadorim.ibict.br/vufind/>. Acesso em 10 jan. 2025.

IBICT. EmeRI. 27 jul. 2023. Disponível em: <https://www.gov.br/ibict/pt-br/itens-capa/produtos-e-servicos/informacao-cientifica/emerri>. Acesso em 10 jan. 2025.

IBICT. BrCris. 10 jan. 2025. Disponível em: <https://brcris.ibict.br/>. Acesso em: 10 jan. 2025.

IBICT. Manuelzão. Brasília, DF, 10 jan. 2025. Disponível em: <https://manuelzao.ibict.br/sobre-o-portal/>. Acesso em: 10 jan. 2025.

IBICT. Miguilim. Brasília, DF, 10 jan. 2025. Disponível em: <https://miguilim.ibict.br/static/pages/miguilim.jsp>. Acesso em: 10 jan. 2025.

IBICT. Deposita Dados: Repositório Comum de Dados de Pesquisa. Brasília, DF, 10 jan. 2025. Disponível: <https://depositadados.ibict.br/about.xhtml>. Acesso em: 10 jan. 2025.

IBICT. Deposita: Repositório Comum do Brasil. Brasília, DF, 20 ago. 2024. Disponível em: <https://www.gov.br/ibict/pt-br/acesso-a-informacao/perguntas-frequentes/sobre-o-deposita>. Acesso em: 10 jan. 2025.

IBICT. Hipátia. Brasilia, DF, 10 jan. 2025. Disponível em: <https://hipatia.ibict.br/o-que-e/>. Acesso em: 10 jan. 2025.

IBICT. Cívis. Brasília, DF, 10 jan. 2025. Disponível em: <https://civis.ibict.br/pt-br/>. Acesso em: 10 jan. 2025.

IBICT. PGD-Br. Brasília, DF, 10 jan. 2025. Disponível: https://pgdbr.ibict.br/?page_id=13. Acesso em: 10 jan. 2025.

IBICT. Rede Moara. Brasilia, DF, 10. Jan. 2025. Disponível em: <https://redemoara.ibict.br/>. Acesso em: 10. jan. 2025.

IBICT. Projeto dARK impulsiona um novo modelo de gestão de identificadores acadêmicos no portal Oasisbr. Brasília, DF, 9 dez. 2024. Disponível em: <https://www.gov.br/ibict/pt-br/central-de-conteudos/noticias/2024/novembro/projeto-dark-impulsiona-um-novo-modelo-de-gestao-de-identificadores-academicos-no-portal-oasisbr-do-ibict>. Acesso em: 10 jan. 2025.

IFLA Statement on Open Access to Scholarly Literature and Research Documentation. **IFLANET**, [Haia], 26 feb. 2004. Disponível em: <https://web.archive.org/web/20040301005941/http://www.ifla.org:80/V/cdoc/open-access04.html>. Acesso em: 10 ago. 2020.

JANKOWSKI, N. W. Exploring E-science: an introduction. **Journal of Computer-Mediated Communication**, [Oxford], v. 12, n. 4, p. 549-562, 2007. Disponível em: <https://onlinelibrary.wiley.com/doi/epdf/10.1111/j.1083-6101.2007.00337.x>. Acesso em: 4 mar. 2021.

JANKOWSKI, N. W. e-Research: transformation in scholarly practice. In: **E RESEARCH: transformation in scholarly practice an enhanced publication accompanying the traditionally published book**. [S. l.]: Routledge, 2009. p. 549-562. Disponível em: <https://academic.oup.com/jcmc/article/12/2/549/4583044>. Acesso em: 4 fev. 2024.

JONES, S; PRYOR, G; WHYTE, A. **How to develop research data management services: a guide for HEIs**. Edimburgo: Digital Curation Centre, 2013. Disponível em: <https://www.dcc.ac.uk/guidance/how-guides/how-develop-rdm-services>. Acesso em 30 jan. 2025.

KUHN, T. S. **A estrutura das revoluções científicas**. São Paulo: Perspectiva, 1975.

KURAMOTO, H. Implantação de repositórios institucionais em universidades e instituições de pesquisa no Brasil: do projeto ao processo. In: ROSA, F.; GOMES, M. J. (org.). **Repositórios institucionais: democratizando o acesso ao conhecimento**. Salvador: Edufba, 2010. p. 61-70.

KURAMOTO, H. Repositórios institucionais: políticas e mandatos. In: SAYÃO, L. et al. **Implantação e gestão de repositórios institucionais: política, memória, livre acesso e preservação**. Salvador: Edufba, 2009. p. 203-218.

LARIVIÈRE, V.; HAUSTEIN, S.; MONGEON, P. The Oligopoly of Academic Publishers in the Digital Era. **PLOS ONE**, San Francisco, p. 1-15, 2015. Disponível em: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0127502>. Acesso em: 12 ago. 2021.

LE COADIC, J. F. **A Ciência da Informação**. Brasília. Briquet de Lemos, 2004.

LEITE, F. C. L. **Como gerenciar e ampliar a visibilidade da informação científica brasileira**: repositórios institucionais de acesso aberto. Brasília, DF: IBICT, 2009. Disponível em: <http://repositorio.unb.br/handle/10482/4841>. Acesso em: 20 jul. 2020.

LICHTMAN, J. W. et al. Discovering the wiring diagram of the brain. In: HEY, T.; TANSLEY, S.; TOLLE, K. (ed.). **The Fourth Paradigm: data intensive scientific discovery**. Redmond: Microsoft Corporation, 2009. p. 83-90. Disponível em: https://www.microsoft.com/en-us/research/uploads/prod/2009/10/Fourth_Paradigm.pdf. Acesso em: 20 abr. 2020.

LIMA, Juliana Soares Lima. **Gestão de Dados de Pesquisa no contexto da Ciência Aberta: percepção dos pesquisadores da Universidade Federal do Ceará**. 2020. 361 f. Dissertação (Mestrado em Ciência da Informação) - Programa de Pós-Graduação em Ciência da Informação, Centro de Humanidades, Universidade Federal do Ceará, Fortaleza, 2020. Disponível em: <http://repositorio.ufc.br/handle/riufc/51572>. Acesso em: 31 mar. 2021.

LUCAS, R.; WILLINSKY, J. Open Access to e-Research. In: JANKOWSKY, N. W. (ed.). **E-research transformation scholarly practice**. New York: Routledge, 2009. Disponível: <https://pkp.sfu.ca/files/e-Research.pdf>. Acesso em: 4 fev. 2022.

LUCE, R. E. E-Science and research libraries: an agenda for action. In: MARCUM, D. B.; GEORGE, G. (ed.). **The data deluge**: can libraries cope with e-science?. Santa Barbara: Libraries Unlimited, 2010. p. 3-16.

LYNCH, C. Jim Gray's fourth paradigm and the construction of the scientific record. In: HEY, T.; TANSLEY, S.; TOLLE, K. (ed.). **The Fourth Paradigm: data intensive scientific discovery**. Redmond: Microsoft Corporation, 2009. p. 177-184. Disponível em: https://www.microsoft.com/en-us/research/uploads/prod/2009/10/Fourth_Paradigm.pdf. Acesso em: 20 abr. 2020.

MARCONDES, C. H.; SAYÃO, L. F. Á guisa de introdução: repositórios institucionais e livre acesso. In: SAYÃO, L. et al. **Implantação e gestão de repositórios institucionais**: política, memória, livre acesso e preservação. Salvador: Edufba, 2009. p. 9-22.

MARCONI, M. de A.; LAKATOS, E. M. **Fundamentos de metodologia científica**. 5. ed. São Paulo: Atlas, 2003.

MÁRDERO-ARELLANO, M. A. Cariniana: uma rede nacional de preservação digital. **Ciência da Informação**, Brasília, DF, v. 41 n. 1, p.83-91, jan./abr., 2012. Disponível em: <https://revista.ibict.br/ciinf/article/view/1354/1533>. Acesso em: 20 abr. 2025.

MARCUM, D. B.; GEORGE, G. (ed.). **The data deluge**: can libraries cope with e-science?. Santa Barbara: Libraries Unlimited, 2010.

MARTÍNEZ-GARCÍA, A. **DOI policy at the University of Cambridge research repository “Apollo”**. [s.l.]: Open Research Systems, Office of Scholarly Communication, 2023. <https://doi.org/10.17863/CAM.10214.2>. Disponível em: <https://www.repository.cam.ac.uk/items/ee597a13-1d26-4f5c-91bb-710edd636d9a>. Acesso em: 20 dez. 2024.

MARTÍNEZ-URIBE, L.; FERNÁNDEZ, P. Servicios de datos: función estratégica de las bibliotecas del siglo XXI. **El Profesional de la información**, Barcelona, v. 24, n. 2, p. 193-199, 2015. Disponível em: <https://revista.profesionaldelainformacion.com/index.php/EPI/article/view/epi.2015.mar.13>. Acesso em: 2 jul. 2024.

MEADOWS, A. J. **A comunicação científica**. Brasília, DF: Briquet de Lemos, 1999.

MERTON, R. K. The normative structure of science. *In: MERTON, R. K. The Sociology of Science*. Chicago: University of Chicago, 1973.

MIKSA, T. et al. **Ten principles for machine-actionable data management plans**. Plos Computational Biology, v. 15, n. 3, e1006750, 2019. DOI: <https://doi.org/10.1371/journal.pcbi.1006750>.

MINISTÉRIO DO ESPORTE. **Classificação dos dados**. Brasília, DF, 17 dez. 2024. Disponível em: <https://www.gov.br/esporte/pt-br/acesso-a-informacao/lgpd/classificacao-dos-dados>. Acesso em: 10 mar. 2025.

MOURA, M. **Ciência na Rua**, [s. l.], 11 mar. 2019. Disponível em: <https://ciencianarua.net/universidades-publicas-respondem-por-mais-de-95-da-producao-cientifica-do-brasil/>. Acesso em: 20 ago. 2024.

MUELLER, S. P. M. A comunicação científica e o movimento de acesso livre ao conhecimento. **Ciência da Informação**, Brasília, DF, v. 35, n. 2, p. 27-38, ago. 2006. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S010019652006000200004&lng=en&nrm=iso. Acesso em: 15 jul. 2020.

NATIONAL SCIENCE FOUNDATION. **Long-Lived Digital Data**. Alexandria Collections: Enabling Research and Education in the 21st Century 2005. [Alexandria]: National Science Fundation, 2005. <https://www.nsf.gov/geo/geo-data-policies/nsb-0540-1.pdf>. Acesso em: 11 set. 2024.

NUNES, M. S. C.; CARVALHO, K. de. As bibliotecas universitárias em perspectiva histórica: a caminho do desenvolvimento durável. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 21, n. 1, p. 173-193, 2016. Disponível em: <https://periodicos.ufmg.br/index.php/pci/article/view/23050>. Acesso em: 13 jun. 2024.

OFFICE OF SCHOLARLY COMMUNICATION; OPEN RESEARCH SYSTEMS. **Repository Terms of Use: Apollo - University of Cambridge Repository**. [S.I.]: Cambridge University Library, 2022. DOI: <https://doi.org/10.17863/CAM.81261>. Disponível em: <https://www.repository.cam.ac.uk/items/a548d870-f37b-4f8c-a038-970b76de6304>. Acesso em: 20 dez. 2024.

OLIVEIRA, A. C. S. de; SILVA, E. M. da. Ciência Aberta: dimensões para um novo fazer científico. **Informação & Informação**, Londrina, v. 21, n. 2, p. 5-39, maio/ago. 2016. Disponível em: https://www.researchgate.net/publication/311853638_Ciencia_aberta_dimensoes_para_um_novo_fazer_cientifico. Acesso em: 2 fev. 2020.

OPEN KNOWLEGDE FUNDATION (OKF). **Open Knowledge Data Tools**. 2024a. Disponível em: <https://okfn.org/en/>. Acesso em: 20. dez. 2024.

OPEN KNOWLEGDE FUNDATION (OKF). **Open Definition**. 2024b. Disponível em: <https://opendefinition.org/guide/>. Acesso em: 20 dez. 2024.

OPEN KNOWLEGDE FUNDATION (OKF). **Open Data Commons. Legal tools for open data**. 21 nov. 2024c. Disponível em: <https://opendatacommons.org/>. Acesso em: 21 nov. 2024.

OPENDOAR. **OpenDOAR Statistics**. [s. /], 2024. Disponível em: https://v2.sherpa.ac.uk/view/repository_visualisations/1.html. Acesso em: 7 jul. 2021.

OPENAIRE. **Guides for Researchers: How to deal with sensitive data**. [s. /], 2024. Disponível em: <https://www.openaire.eu/sensitive-data-guide>. Acesso em: 20 dez. 2024.

OPENAIRE. **GDPR and the research process: What you need to know**. [s. /], 2018. Disponível: <https://www.openaire.eu/blogs/gdpr-and-the-research-process-what-you-need-to-know?Itemid=2068&highlight=WyJzZW5zaXRpdmUiLCJkYXRhliwiZGF0YSdzliwiJ2RhGEiLCJkYXRhJylsInNlbnNpdGI2ZSBkYXRhIi0=>. Acesso em: 15 abr. 2025.

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT. **Principles and Guidelines for Access to Research data from Public Funding**. Paris: OECD, 2007. Disponível em: https://www.oecd-ilibrary.org/science-and-technology/oecd-principles-and-guidelines-for-access-to-research-data-from-public-funding_9789264034020-en-fr. Acesso em: 2 ago. 2020.

PARCERIA PARA GOVERNO ABERTO. **4º plano de ação nacional em Governo Aberto**. Brasília: Controladoria Geral da União (CGU), 2018. Disponível em: <https://repositorio.cgu.gov.br/handle/1/66740>. Acesso em: 26 abr. 2021.

PARCERIA PARA GOVERNO ABERTO. **5º plano de ação nacional em Governo Aberto**. Brasília: Controladoria Geral da União (CGU), 2021. Disponível em: <https://repositorio.cgu.gov.br/handle/1/67649> Acesso em: 15 jan. 2024.

PARCERIA PARA GOVERNO ABERTO. **6º plano de ação nacional em Governo Aberto**. Brasília: Controladoria Geral da União (CGU), 2023. Disponível em: https://www.gov.br/cgu/pt-br/governo-aberto/a-ogg/planos-de-acao/6deg-plano-de-acao-brasileiro/brazil_action-plan_2023-2027_december_pt.pdf. Acesso em: 15 jan. 2024.

PAVÃO, C. G.; ROCHA, R. P.; GABRIEL JUNIOR, R. F. Proposta de criação de uma rede de dados abertos da pesquisa brasileira. **RDBCi: Revista Digital de Biblioteconomia e Ciência da Informação, Campinas**, SP, v. 16, n. 2, p. 329–343, 2018. DOI: 10.20396/rdbcii.v16i2.8651180. Disponível em: <https://periodicos.sbu.unicamp.br/ojs/index.php/rdbcii/article/view/8651180>. Acesso em: 10 jan. 2025.

PEREIRA, V.; FURNIVAL, A. C. Revistas científicas em acesso aberto brasileiras no DOAJ: Modelos de negócio e sua sustentabilidade financeira. **Brazilian Journal of Information Science: research trends**, Marília, v. 14, n. 1, p. 88-111, jan. /mar. 2020. Disponível em: <https://revistas.marilia.unesp.br/index.php/bjis/article/view/9695>. Acesso em: 3 jul. 2024.

PLALE, B.; KOUPER, I. The Centrality of Data: Data Lifecycle and Data Pipelines. In: CHOWDRY, M.; APON, A.; DEY, K. **Data Analytics for Intelligent Transportation Systems**. Amsterdam: Elsevier, 2017. p. 91-111. Disponível em: https://www.researchgate.net/publication/315861215_The_Centrality_of_Data_Data_Lifecycle_and_Data_Pipelines. Acesso em: 4 jul. 2024.

PONTIKA, N. et al. Fostering open science to research using a taxonomy and an eLearning portal. In: **IKNOW: 15th International Conference on Knowledge Technologies and Data Driven Business**, 21-22 october 2015, Graz, Austria, 2015. Disponível em: <http://oro.open.ac.uk/44719/>. Acesso em: 21 nov. 2024.

PORTAL CIÊNCIA DE ABERTA. **Ciência Aberta – uma abordagem colaborativa para o processo científico**. 2025. Disponível em: <https://cienciaaberta.org/ciencia-aberta/> Acesso em: 12 jan. 2025.

PRÍNCIPE, P. et al. **Relatório técnico sobre ferramentas para a elaboração de Planos de Gestão de Dados**. [S. I.]: ACAAP, 2020. Disponível em: https://repositorium.sdum.uminho.pt/bitstream/1822/67530/1/RelatorioTecnico_Ferramentas_PlanosGestao_DadosV1_final_23092020.pdf. Acesso em: 20 jan. 2025.

PROJECT GUTEMBERG. About Project Gutenberg. **Project Gutenberg**, [s. l.], [202-]. Disponível em: <https://www.gutenberg.org/about/>. Acesso em: 8 jul. 2021.

PUBLIC KNOWLEDGE PROJECT. About. **Public Knowledge Project**, [s. l.], 2021. Disponível em: <https://pkp.sfu.ca/>. Acesso em: 10 dez. 2023.

QUINTANILHA, T. L. Considerações sobre os desafios múltiplos da Ciência Aberta. **Estudos em Comunicação**, Lisboa, n. 21, p. 13-24, dez. 2015. Disponível em: https://repositorio.iscte-iul.pt/bitstream/10071/10563/5/ec_21_02.pdf. Acesso em: 28 fev. 2024.

REDE BRASILEIRA PARA EDUCAÇÃO E PESQUISA. **Rede de e-Ciência conecta os grandes projetos de pesquisa no Brasil**. Brasília, 21 jan. 2025. Disponível em: <https://www.rnp.br/rede-de-e-ciencia-conecta-os-grandes-projetos-de-pesquisa-no-brasil/>. Acesso em: 20 mar. 2025.

REDE BRASILEIRA PARA EDUCAÇÃO E PESQUISA (RNP). **Alteryx Analytics**. 20 mar. 2025. Disponível em: <https://rnpmais.rnp.br/alteryx#adquirir>. Acesso em: 20 mar. 2025.

REDE BRASILEIRA PARA EDUCAÇÃO E PESQUISA (RNP). **Sistema de Informação sobre a Biodiversidade Brasileira**. 26 jan. 2025. Disponível em: <https://sibbr.gov.br/page/o-que-sibbr.html> Acesso em: 26 jan. 2025.

REDE DE DADOS DE PESQUISA BRASILEIRA (RNP). **RDP Brasil**. 20 dez. 2024. Disponível em: <https://www.ufrgs.br/rdpbrasil/wp/> Acesso em: 20 dez. 2024.

REINSEL, D.; GANTZ, J.; RYDNING, J. **Data age 2025**: the evolution of data to life-critical (don't focus on Big Data, focus on the Data that's big). New York: IDC, 2017. Disponível em: <https://www.seagate.com/www-content/our-story/trends/files/Seagate-WP-DataAge2025-March-2017.pdf>. Acesso em: 21 set. 2021.

RE3DATA. **Countries**. 2025. Disponível em: <https://www.re3data.org/search?query=&countries%5B0%5D=GBR&page=2>. Acesso em: 20 mar. 2025.

RESEARCH DATA MANAGEMENT LIBRARIAN ACADEMY (RDMLA). **Course RDMLA**. 2025. Disponível em: <https://rdmla.github.io/> Acesso em: 16 fev. 2025.

RIOS, F. P.; LUCAS, E. R. O.; AMORIM, I. S. Manifestos do movimento de acesso aberto: Análise de Domínio a partir de periódicos brasileiros. **Revista Brasileira de Biblioteconomia e Documentação**, São Paulo, v. 15, n. 1, p. 148-169, 2019. Disponível em: <https://rbbd.febab.org.br/rbbd/article/view/1152>. Acesso em: 21 set. 2021.

ROCHA, R. P.; GABRIEL JUNIOR, R. F.; VANZ, S. A. de S.; BORGES, E. N.; AZAMBUJA, L. A. B.; CAREGNATO, S. E.; PAVÃO, C. G.; PASSOS, P. C. S. J.; FELICISSIMO, C. H. Análise dos sistemas DSpace e Dataverse para repositórios de dados de pesquisa com acesso aberto. **Revista Brasileira de Biblioteconomia e Documentação**, [S. I.], v. 17, p. 1–25, 2021. Disponível em: <https://rbbd.febab.org.br/rbbd/article/view/1572>. Acesso em: 20 jan. 2025.

ROSA, F.; GOMES, M. J. Comunicação científica: das restrições ao acesso livre. In: ROSA, F.; GOMES, M. J. (org.). **Repositórios institucionais**: democratizando o acesso ao conhecimento. Salvador: Edufba, 2010. p. 153-204.

RUDIO, F. V. **Introdução ao projeto de pesquisa científica**. Petrópolis: Vozes, 2007.

RUUSALEPP, R. **A comparative study of the international approaches to enabling the sharing of research data**. [S. I.]: Digital Data Curation, 2008. Disponível em: <https://era.ed.ac.uk/handle/1842/3361>. Acesso em: 20 jul. 2024.

SALES, L. F. et al. Competências dos bibliotecários na gestão dos dados de pesquisa. **Ciência da Informação**, Brasília, DF, v. 48, n. 3, p. 303-313, set./dez. 2019. Supl. Disponível em: <https://revista.ibict.br/ciinf/article/view/4973/4458>. Acesso em: 20 jul. 2024.

SALES, L. F. **Integração semântica de publicações científicas e dados de pesquisa**: proposta de modelo de publicação ampliada para a área de ciências nucleares. 2014. Tese (Doutorado em Ciência da Informação) – Instituto Brasileiro de Informação e Tecnologia, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2014. Disponível em: <https://ridi.ibict.br/handle/123456789/874>. Acesso em: 21 jun. 2024.

SALVADOR. **Declaration on Open Access**: The Developing World Perspective. Salvador: ICML, 2005. Disponível em: <http://www.icml9.org/meetings/openaccess/public/documents/declaration.htm>. Acesso em: 11 ago. 2020.

SANTA FE. **The Santa Fe Convention for the Open Archives Initiative**. Santa Fe: [s. n.], 2000. Disponível em: https://www.openarchives.org/sfc/sfc_entry.htm. Acesso em: 2 dez 2023.

SANTOS, P. X. et., al (Coord.). **Livro Verde - Ciência aberta e dados abertos: mapeamento e análise de políticas, infraestruturas e estratégias em perspectiva nacional e internacional**. Rio de Janeiro: Fiocruz, 2017. Disponível em: <https://www.arca.fiocruz.br/handle/icict/24117>. Acesso em: 22 jul. 2021.

SARACEVIC, T. **Information Science**. Journal of the American Society for Information Science (JASIS), v. 59, n. 12, p. 1051-1063, 1999. Disponível em: [https://asistdl.onlinelibrary.wiley.com/doi/10.1002/\(SICI\)1097-4571\(1999\)50:12%3C1051::AID-ASI2%3E3.0.CO;2-Z](https://asistdl.onlinelibrary.wiley.com/doi/10.1002/(SICI)1097-4571(1999)50:12%3C1051::AID-ASI2%3E3.0.CO;2-Z). Acesso em: 20 out. 2024.

SAYÃO, L. F.; SALES, L. F. A ciência invisível: os dados da cauda longa da pesquisa. In: DIAS, G. A.; OLIVEIRA, B. M. J. F. de (org.). **Dados científicos: perspectivas e desafios**. João Pessoa: UFPB, 2019. p. 33-52.

SAYÃO, L. F.; SALES, L. F. Algumas considerações sobre os repositórios digitais de dados de pesquisa. **Informação & Informação**, Londrina, v. 21, n. 2, p. 90-115, dez. 2016. Disponível em: <http://www.uel.br/revistas/uel/index.php/informacao/article/view/27939/20122>. Acesso em: 15 ago. 2020.

SAYÃO, L. F.; SALES, L. F. Afinal, o que é dado de pesquisa. **Biblos**, Rio Grande, v. 34, n. 2, p. 32-51, jul./dez. 2020. Disponível em: <https://periodicos.furg.br/biblos/article/download/11875/8426/38>. Acesso em: 20 jul. 2021.

SAYÃO, L. F.; SALES, L. F. **Guia de gestão de dados de pesquisa para bibliotecários e pesquisadores**. Rio de Janeiro: CNEN/IEN, 2015. Disponível em: <http://www.cnen.gov.br/component/content/article/75-cin/material-didatico-cnen/160-guia-de-gestao-de-dados-de-pesquisa>. Acesso em: 10 jul. 2020.

SAYÃO, L. F.; SALES, L. F. Plataformas de gestão de dados de pesquisa: expandindo o conceito de repositórios de dados. **Palabra Clave**, La Plata, v. 12, nº 1, p. 1-21, out/mar. 2022-2023. DOI: <https://doi.org/10.24215/18539912e171>.

Disponível em: <https://www.palabraclave.fahce.unlp.edu.ar/article/view/pce171>. Acesso em: 10 abr. 2024.

SCHNELL, Santiago. Ten Simple Rules for a Computational Biologist's Laboratory Notebook. **PLoS Computational Biology**, São Francisco, v. 11, n. 9, 2015.

Disponível em:

<https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1004385>.

Acesso em: 6 jan. 2025.

SECCHI, L. **Análise de políticas públicas**: diagnóstico de problemas, recomendação de soluções. São Paulo: Cengage, 2016.

SEMELER, A. R.; PINTO, A. L. Os diferentes conceitos de dados de pesquisa na abordagem da biblioteconomia de dados. **Ciência da Informação**, Brasília, DF, v. 48 n.1, p.130-129, jan./abr. 2019. Disponível em:

<http://revista.ibict.br/ciinf/article/view/4461/4102>. Acesso em: 25 jul. 2020.

SEMELER, A. R.; PINTO, A. L. Data librarianship as a field study. **Transinformação**, [S. I.], v. 32, 2020. Disponível em: <https://periodicos.puc-campinas.edu.br/transinfo/article/view/5851>. Acesso em: 8 abr. 2025.

SEMIDÃO, R. A. M. **Dados, Informação e Conhecimento enquanto elementos de compreensão do universo conceitual da Ciência da Informação**: contribuições teóricas. 2014. Dissertação (Mestrado em Ciência da Informação) –Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2014. Disponível em: <https://repositorio.unesp.br/bitstream/handle/11449/110783/000799485.pdf?sequencia=1>. Acesso em: 23 out. 2023.

SEVERINO, A. J. **Metodologia do trabalho científico**. São Paulo: Cortez, 2013.

Disponível em:

https://www.ufrb.edu.br/ccaab/images/AEPE/Divulga%C3%A7%C3%A3o/LIVROS/Metodologia_do_Trabalho_Cient%C3%ADfico_-_1%C2%AA_Edi%C3%A7%C3%A3o_-_Antonio_Joaquim_Severino_-_2014.pdf. Acesso em: 5 jul. 2024.

SHARMA, N. **The origin of Data Information Knowledge Wisdom (DIKW) hierarchy**. [S. I.: s. n.], 2008. Disponível em:

https://www.researchgate.net/publication/292335202_The_Origin_of_Data_Information_Knowledge_Wisdom_DIKW_Hierarchy. Acesso em: 21 jan. 2021.

SHINTAKU, M.; FERREIRA JÚNIOR, R. S. Política editorial e os dados abertos. In: SHINTAKU, M.; SALES, L. F.; COSTA, M. (org.). **Tópicos sobre dados abertos para editores científicos**. Botucatu: ABEC, 2020. p. 231-240.

SILVA, F.C. C. da. O Papel dos bibliotecários na gestão de dados científicos.

RDBCi: Revista Digital de Biblioteconomia e Ciência da Informação, Campinas, SP, v. 14, n. 3, p. 387-406, set/dez. 2016. DOI: 10.20396/rdbcii.v14i3.8646333.

Disponível em:

<https://periodicos.sbu.unicamp.br/ojs/index.php/rdbcii/article/view/8646333>. Acesso em: 20 dez. 2023.

SILVA, F. C. C. da. **Gestão de dados científicos**. Rio de Janeiro: Interciência, 2019.

SILVA, F. C. C. da; SILVEIRA, L. da. O ecossistema da Ciência Aberta. **Transinformação**, Campinas, v. 31, 2019. Disponível em: <https://periodicos.puc-campinas.edu.br/transinfo/article/view/5926>. Acesso em: 16 nov. 2024.

SILVA, F. C. C da.; SCHWAITZER, L. de B. da S.; REZENDE, L. V. R. **Análise terminológica de dados de pesquisa científica para preservação digital**. In: ARELLANO, M. A. M.; ESTRELA, F. M. S.; SILVA, A. S. Temas de pesquisa em preservação digital. Brasília, DF: IBICT, 2024. p. 195-219. Disponível em: <https://omp-editora.prd.ibict.br/index.php/edibict/catalog/view/339/376/2178>. Acesso em: 10 nov. 2024.

SILVEIRA, D. T.; CÓRDOVA, F. P. A pesquisa científica. In: GERHARDT, T. E.; SILVEIRA, D. T. (org.). **Métodos da pesquisa**. Porto Alegre: Ed. UFRGS, 2009. p. 31-42.

SILVEIRA, L. et al. Ciência aberta na perspectiva de especialistas brasileiros: proposta de taxonomia. **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação**, [S. I.], v. 26, p. 1–27, 2021. DOI: 10.5007/1518-2924.2021.e79646. Disponível em: <https://periodicos.ufsc.br/index.php/eb/article/view/79646>. Acesso em: 19 dez. 2024.

SILVEIRA, L. et al. Taxonomia da Ciência Aberta: revisada e ampliada. **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação**, [S. I.], v. 28, p. 1–22, 2023. DOI: 10.5007/1518-2924.2023.e91712. Disponível em: <https://periodicos.ufsc.br/index.php/eb/article/view/91712>. Acesso em: 19 dez. 2024

SINAEEPOURFARD, A. et al. Towards a Comprehensive Data LifeCycle Model for Big Data Environments. In: IEEE/ACM INTERNATIONAL CONFERENCE ON BIG DATA COMPUTING APPLICATIONS AND TECHNOLOGIES, 3., 2016, Sharjah. **Anais** [...]. Sharjah: [s. n.], 2016. 100-106. Disponível em: <https://www.semanticscholar.org/paper/Towards-a-Comprehensive-Data-LifeCycle-Model-for-Sinaeepourfard-Garc%C3%ADa5d01b138b1600469208ef4a1c6ee821dfafcd6>. Acesso em: 5 jul. 2024.

SMALE, N. et al. The history, advocacy and efficacy of data management plans. **bioRxiv**, 2018. Doi: <https://doi.org/10.1101/443499>. Disponível em: <https://www.biorxiv.org/content/10.1101/443499v1> Acesso em 15 jan. 2025.

SMITH, C. **Cambridge University Libraries Digital Preservation Policy (V2.0)**. [S. I.]: Cambridge University Libraries, 2021. <https://doi.org/10.17863/CAM.71114>. Disponível em: <https://www.repository.cam.ac.uk/items/dc4deb94-e6db-4ff7-99eb-9d1e27340393>. Acesso em: 20 dez. 2024.

SOCIENTIZE CONSORTIUM. **Green Paper on Citizen Science: citizen science for Europe towards a better society of empowered citizens and enhanced research**. Bruxelas: European Comission, 2013. Disponível em: <https://digital>

strategy.ec.europa.eu/en/library/green-paper-citizen-science-europe-towards-society-empowered-citizens-and-enhanced-research.
Acesso em: 27 jul. 2024.

SOLON, O. Facebook says Cambridge Analytica may have gained 37m more users' data. **The Guardian**, Reino Unido, 2018. Disponível em: <https://www.theguardian.com/technology/2018/apr/04/facebook-cambridge-analytica-user-data-latest-more-than-thought>. Acesso em: 27 fev. 2024.

SOUZA, M. de; ALMEIDA, F. G. O Comportamento do termo dado na Ciência da Informação. **Ciência da Informação em Revista**, Maceió, v. 8, n. 2, p. 39-54, 2021. Disponível em: <https://www.seer.ufal.br/index.php/cir/article/view/11764>. Acesso em: 10 jun. 2024.

SPICHTINGER, D. Data Management Plan in Horizon 2020: what beneficiaries think and what we can learn from their experience. **Open Research Europe**, Viena, v. 1, n. 42, p. 1-22, 2022. DOI: <https://doi.org/10.12688/openreseurope.13342.2>
Disponível em: <https://open-research-europe.ec.europa.eu/articles/1-42>. Acesso em: 23 nov. 2024.

STEELEWORTHY, M. Research Data Management and the Canadian Academic Library: An Organizational Consideration of Data Management and Data Stewardship. **Partnership The Canadian Journal of Library and Information**, [s. l.], v. 9, n. 1, p. 1-11, 2014. Disponível em: https://www.researchgate.net/publication/263309264_Research_Data_Management_and_the_Canadian_Academic_Library_An_Organizational_Consideration_of_Data_Management_and_Data_Stewardship. Acesso em: 13 jun. 2024.

SUBER, P. et al. **Bethesda Statement on Open-Access Publishing**. Dash, Cambridge, [200-]. Disponível em: https://www.researchgate.net/publication/48547523_The_Bethesda_Statement_on_Open-Access_Publishing. Acesso em: 20 jul. 2020.

SUBER, P. Open Access. Massachusetts: MIT Press, 2012. Disponível em: <https://direct.mit.edu/books/book/3754/Open-Access>.
Acesso em: 20 jul. 2020.

SYRACUSE UNIVERSITY. **The Qualitative Data Repository**. 2025. Disponível em: <https://qdr.syr.edu/>. Acesso em: 16 fev. 2025.

TANG, R.; HU, Z. Providing Research Data Management (RDM) Services in Libraries: Preparedness, Roles, Challenges, and Training for RDM Practice. **Data and Information Management**, [S. l.], v. 3, n. 2, p. 84-101, 2019. Disponível em: <https://www.sciencedirect.com/science/article/pii/S2543925122000675#bib026>
Acesso em: 20 dez. 2024.

TAURION, C. **Big Data**. Rio de Janeiro: Brasport, 2013. Disponível em: <https://doceru.com/doc/sn0xe5e>. Acesso em: 25 set. 2021.

TENOPIR, C. *et al.* Research data services in european and north American libraries: Current offerings and plans for the future. **Proceedings of the Association for Information Science and Technology**, [S. I.], v. 53, n. 1, p. 1–6, 2016. Disponível em: <http://doi.wiley.com/10.1002/pra2.2016.14505301129>. Acesso em: 25 jul. 2024.

UNIVERSITY OF MICHIGAN. **The Inter-university Consortium for Political and Social Research (ICPSR)**. 2025. Disponível em: <https://www.icpsr.umich.edu/web/pages/index.html>. Acesso em: 16 fev. 2025.

TILMAN, D. *et al.* Agricultural sustainability and intensive production practices. **Nature**, London, v. 418, p. 671-677, 2002. Disponível em: https://www.researchgate.net/publication/232770232_review_articleAgricultural_sustainability_and_intensive_production_practices. Acesso em: 21 jan. 2024.

UNESCO. **Recomendação da UNESCO sobre Ciência Aberta**. [S. I.]: Unesco, 2022. Disponível em: https://unesdoc.unesco.org/ark:/48223/pf0000379949_por. Acesso em: 2 dez. 2023.

UNIVERSIDADE FEDERAL DO CEARÁ. **Repositório de Dados de Pesquisa Antônio Martins Filho**. 2025. Disponível em: <https://dadosdepesquisa.ufc.br/>. Acesso em: 30 abr. 2025.

UNIVERSIDADE FEDERAL DO PARANÁ. Biblioteca Digital de Periódicos. Sobre o OBJS. **Revistas UFPR**, Curitiba, 2023. Disponível em: [https://revistas.ufpr.br/wp/sobre-o-ojs/#:~:text=O%20Open%20Journal%20Systems%20\(OJS,de%20Stanford%2C%20sede%20do%20projeto](https://revistas.ufpr.br/wp/sobre-o-ojs/#:~:text=O%20Open%20Journal%20Systems%20(OJS,de%20Stanford%2C%20sede%20do%20projeto). Acesso em: 20 set. 2023.

UNIVERSITEIT LEIDEN. Open Science. **Universiteit Leiden**, Leiden 2024. Disponível em: <https://www.library.universiteitleiden.nl/researchers/open-science>. Acesso em: 11 set. 2024.

UNIVERSITY OF SHEFFIELD. Department of Computer Science. Clive Humby. **University of Sheffield**, Sheffield, c2024. Disponível em: <https://www.sheffield.ac.uk/dcs/people/academic-visitors/clive-humby>. Acesso em: 26 fev. 2024.

VALENTE, M. G.; HOUANG, A. **Creative Commons Br: o que você precisa saber sobre licenças CC**. [s./.]: Creative Commons Brasil, 2020. Disponível em: <https://br.creativecommons.net/wp-content/uploads/sites/30/2021/02/CartilhaCCBrasil.pdf>. Acesso em: 21 nov. 2024.

VALPARAÍSO. **Valparaíso declaration for improved scientific communication in the electronic medium**. Valparaíso: [s. n.], 2004. Disponível em: https://www.academia.edu/79562763/Valpara%C3%ADso_declaration_for_improved_scientific_communication_in_the_electronic_medium. Acesso em: 8 jul. 2021.

VAN DE SOMPEL, H.; LAGOZE, C. The Santa Fe Convention of the Open Archives Initiative. **D-Lib Magazine**, [s. I.], v. 6, n. 2, fev. 2000. Disponível em:

<https://www.dlib.org/dlib/february00/vandesompel-oai/02vandesompel-oai.html>. Acesso em: 26 set. 2021.

VEIGA, V. et al. VODAN BR: a gestão de dados no enfrentamento da pandemia coronavírus. **Páginas a&b**, Porto, v. 3, p. 51-58, 2021. Número especial. Disponível em: <https://www.arca.fiocruz.br/handle/icict/46443>. Acesso em: 21 abr. 2024.

VEIGA, V. et al. **Plano de gestão de dados de pesquisa: guia de elaboração**. Rio de Janeiro: FIOCRUZ, ICICT, Rede de Bibliotecas, 2022. Disponível em: <https://sddinforma.fob.usp.br/plano-de-gestao-de-dados-de-pesquisa-pgd-guia-de-elaboracao/>. Acesso em: 13 abr. 2023.

VICENT-SAEZ, R.; MARTINEZ-FUENTES, C. Open Science now: A systematic literature review for an integrated definition. **Journal of Business Research**, [s. L.], v. 88, p. 428-436, jul. 2018. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0148296317305441?via%3Dhub> Acesso em: 20 dez. 2024.

Vidal, L.H.C.; Gabriel Junior, R.F.; Pavão, C.M.G. Ferramentas para elaboração de planos de gestão de dados: visão geral e análise. **Transinformação**, v. 37, 2025. Disponível em: <https://doi.org/10.1590/2318-0889202537e2510891>. Acesso em: 20 set. 2024.

VIENA. **The Vienna Declaration**. Viena: [s. n.], 1993. Disponível em: https://web.archive.org/web/20060304153832/http://www.chaoscontrol.at:80/2005/we_english.htm. Acesso em: 10 ago. 2020.

WILKINSON, M. D. et. al. The FAIR guiding Principles for scientific data management and stewardship. **Sci Data**, [s. l.], v. 3, 2016. Disponível em: <https://www.nature.com/articles/sdata201618>. Acesso em: 19 nov. 2024.

WHITMIRE, A. L. eResearch: a scientist's perspective. **Journal of eScience Librarianship**, Massachusetts, v. 2, n. 2, jul. 2013. Disponível em: https://www.academia.edu/9097345/Thoughts_on_eResearch_a_Scientists_Perspective?auto=download. Acesso em: 15 abr. 2020.

WOUTERS, P.; BEAULIEU, A. Critical Accountability: Dilemmas for Interventionist Studies of e-Science. **Journal of Computer-Mediated Communication**, Oxford, v. 12, n. 2, p. 583-599, jan. 2007. Disponível em: <https://onlinelibrary.wiley.com/doi/full/10.1111/j.1083-6101.2007.00339.x>. Acesso em: 15 jul. 2021.

ZIMAN, J. **Conhecimento público**. São Paulo: Edusp, 1979.

APÊNDICE A – QUESTIONÁRIO

Este questionário é um instrumento para pesquisa de campo e visa compreender como as bibliotecas das universidades federais do Nordeste brasileiro estão se preparando para promover o suporte à gestão dos dados científicos da universidade.

* Indica uma pergunta obrigatória

1 - Qual o nome da universidade que o sistema de bibliotecas é vinculado? *

Universidade Federal de Alagoas (UFAL)

Universidade Federal da Bahia (UFBA)

Universidade Federal do Recôncavo da Bahia (UFRB)

Universidade da Integração Internacional da Lusofonia Afro-Brasileira (UNILAB)

Universidade Federal do Ceará (UFC)

Universidade Federal do Maranhão (UFMA)

Universidade Federal da Paraíba (UFPB)

Universidade Federal de Campina Grande (UFCG)

Universidade Federal de Pernambuco (UFPE)

Universidade Federal Rural de Pernambuco (UFRPE)

Universidade Federal do Piauí (UFPI)

Universidade Federal do Rio Grande do Norte (UFRN)

Universidade Federal do Semi-Árido (UFERSA)

Universidade Federal de Sergipe (UFS)

2 - O sistema de bibliotecas possui Política/Norma implementada para gestão dos dados científicos dos pesquisadores da universidade? *

a) Possui.

b) Não possui, mas prevê planejamento para sua formulação.

c) Não possui e nem prevê planejamento para sua formulação.

2.1 – Caso tenha respondido “possui” na pergunta anterior, por favor inserir o link da Política/Norma a abaixo ou envie o pdf para email uillisassis@hotmail.com?

3 - O sistema de bibliotecas divulga informações sobre gestão de dados científicos através do seu site, das suas redes sociais ou e-mail institucional? *

- a) Divulga.
- b) Não divulga, mas prevê planejamento para sua divulgação.
- c) Não divulga e nem prevê planejamento para sua divulgação.

3.1 – Caso tenha respondido “Divulga” na pergunta anterior, por favor insira o link de pelo menos uma divulgação.

4 – O sistema de bibliotecas promove ações de sensibilização sobre a importância do gerenciamento dos dados científicos para comunidade da universidade através de palestras? *

- a) Promove.
- b) Não promove, mas prevê planejamento para sua constituição.
- c) Não promove e nem prevê planejamento para sua constituição.

4.1 - Caso tenha respondido "Promove" na pergunta anterior, por favor insira informações básicas sobre o conteúdo das palestras.

5 – O sistema de bibliotecas disponibiliza suporte financeiro para participação de bibliotecários (a) em cursos de capacitação em gestão de dados científicos? *

- a) Disponibiliza.
- b) Não disponibiliza, mas prevê planejamento para sua divulgação.
- c) Não disponibiliza e nem prevê planejamento para sua divulgação.

5.1 - Caso tenha respondido "Disponibiliza" na pergunta anterior, por favor informe o nome do (s) curso (s) que a equipe já participou.

6 – O sistema de bibliotecas possui serviço de orientação sobre o uso dos Princípios FAIR? *

- a) Possui.
- b) Não possui, mas prevê planejamento para sua constituição.
- c) Não possui e nem prevê planejamento para sua constituição.

6.1 - Caso tenha respondido "Possui" na pergunta anterior, por favor descreva o serviço.

7 – O sistema de bibliotecas possui serviço de orientação sobre o funcionamento do Ciclo de Vida dos Dados na gestão de dados científicos? *

- a) Possui.
- b) Não possui, mas prevê planejamento para a formulação do incentivo.
- c) Não possui e nem prevê planejamento para a formulação do incentivo.

7.1 - Caso tenha respondido "Possui" na pergunta anterior, por favor descreva o serviço.

8 – O sistema de bibliotecas possui serviço de orientação para o preenchimento do Plano de Gestão de Dados (PGD) das pesquisas desenvolvidas na universidade? *

- a) Possui.
- b) Não possui, mas prevê planejamento para a formulação.
- c) Não possui e nem prevê planejamento para a formulação.

8.1 - Caso tenha respondido "Possui" na pergunta anterior, por favor descreva o serviço.

9 – O sistema de bibliotecas possui serviço de orientação para anonimização dos dados pessoais e sensíveis dos conjuntos de dados pertencentes as pesquisas desenvolvidas na universidade? *

- a) Possui.
- b) Não possui, mas prevê planejamento para sua constituição.
- c) Não possui e nem prevê planejamento para sua constituição.

9.1 - Caso tenha respondido "Possui" na pergunta anterior, por favor descreva o serviço.

10 – O sistema de bibliotecas possui serviço de orientação para os pesquisadores da universidade encontrarem dados úteis para o desenvolvimento de suas pesquisas? *

- a) Possui.
- b) Não possui, mas prevê planejamento para sua constituição.
- c) Não possui e nem prevê planejamento para sua constituição.

10.1 - Caso tenha respondido "Possui" na pergunta anterior, por favor descreva o serviço.

11 – A universidade possui repositório de dados científicos? *

Entende-se nessa pesquisa:

Repositórios de publicação é um serviço de informação científica que possibilita o gerenciamento e a disseminação da produção acadêmica da Universidade (teses, dissertações, artigos científicos, artigo de evento, livros e capítulos de livros. (Leite et al., 2012)

Repositórios de dados científicos é um serviço de informação científica que possibilita o gerenciamento e a disseminação dos dados científicos gerados nas pesquisas científicas. Os dados científicos são constituídos por uma tipologia variada: números, questionários, imagens, avaliações, textos, registros de casos, vídeos, áudio, protocolos de estudo, software, banco de dados, planilhas, algoritmos, notas de laboratório, equações, notas de campo, animações, sequências genéticas, modelos, metodologias, protocolos, diários, simulações e coordenadas geográficas (Sayão; Sales, 2015; Semeler; Pinto, 2019; Silva, 2019; Paganine; Amaro, 2020)

- a) Possui.
- b) Não possui, mas prevê planejamento para sua constituição.
- c) Não possui e nem prevê planejamento para sua constituição.

11.1 - Caso tenha respondido "Possui" na pergunta anterior, por favor descreva a função do repositório de dados científicos.

12 – O sistema de bibliotecas possui serviço de orientação para indicar repositórios de dados científicos para o depósito dos conjuntos de dados dos pesquisadores da universidade? *

- a) Possui.
- b) Não possui, mas prevê planejamento para sua constituição.
- c) Não possui e nem prevê planejamento para sua constituição.

12.1 - Caso tenha respondido "Possui" na pergunta anterior, por favor descreva o serviço.

13 – Descreva neste espaço caso o sistema de bibliotecas possua algum serviço voltado para o gerenciamento de dados científicos ou planejamento para este propósito além dos previstos neste questionário.

- a) Possui.
- b) Não possui, mas prevê planejamento para sua participação.
- c) Não possui e nem prevê planejamento para sua participação.

APÊNDICE B – PLANO DE GESTÃO DE DADOS DA DISSERTAÇÃO

Gestão de dados científicos nas bibliotecas das universidades federais do Nordeste brasileiro: análise à luz da Ciência Aberta

Iniciar o plano

Qual modelo vai utilizar?

- Plano de Gestão de Dados - Template resumido

Coleta de Dados

Os dados serão:

- Coletados ou produzidos por mim ou minha equipe

Os dados foram produzidos pelo responsável pela pesquisa.

Do ponto de vista de sua natureza, quais tipos de dados serão coletados, produzidos ou reutilizados? Assinale quantas opções forem necessárias.

- Número (ex.: medidas, resultado de levantamento, fórmula, equação, algoritmo)
- Visualização (ex.: tabelas, gráficos, diagramas, modelo em 3D, modelo reduzido, desenho)
- Textual (ex.: metadado, questionário, entrevista, anotação, certificado, caderno de laboratório, transcrição, correspondência, diário, caderno de campo)
- Processo (ex.: procedimentos operacionais padronizados, workflow, protocolo, teste)

Quais os formatos dos dados que serão coletados, produzidos ou reutilizados?

- PDF/A - Portable Document Format

Qual o volume aproximado dos dados coletados, produzidos ou reutilizados?

- 0 a 1GB

Descreva como os dados serão coletados ou criados?

Os dados serão coletados através de questionários do google forms enviados para as coordenações e diretorias dos sistemas de bibliotecas das universidades federais do Nordeste

brasileiro.

Seu conjunto de dados contém dados pessoais ou sensíveis?

- Não

Documentação e Metadados

Qual documentação irá acompanhar os dados?

- Nenhuma

Qual padrão de metadados será utilizado?

Padrão de metadados do DATAVERSE

Armazenamento, Backup e Preservação

Durante a pesquisa, onde o(s) conjunto(s) de dados será(ão) armazenado(s)?

- Armazenamento nas Nuvens Privado

Ao finalizar a pesquisa, onde o(s) conjunto(s) de dados serão armazenados (armazenamento definitivo)?

- Repositório de dados temáticos ou gerais disponíveis para toda a comunidade científica (Ex: Zenodo, Figshare, Harvard Dataverse);
- Armazenamento nas Nuvens Privado;

Como será feito o backup dos dados?

- Backup realizado por serviços de terceiros (Ex: Google Backup);

Durante a coleta, os dados e a documentação serão armazenados em computador pessoal e em nuvem com backup em ambos a cada atualização na coleta, edição ou análise

Como será gerenciado o acesso aos dados durante a pesquisa?

- A equipe de pesquisa usará um único login e senha no sistema para todos da equipe de pesquisa manipularem os dados;

Como será gerenciado o acesso aos dados ao final da pesquisa?

- A equipe de pesquisa usará um único login e senha no sistema para todos da equipe de pesquisa caso ainda precise manipular os dados;
- O acesso aos dados será gerenciado pelo repositório onde estarão inseridos;

Durante a pesquisa, se estiver criando ou coletando dados em campo, como você garantirá a transferência segura para seus principais sistemas?

- Membros da equipe realizam a coleta de dados em campo utilizando dispositivos móveis COM acesso à Internet; Durante ou imediatamente ao final de cada etapa, os dados são inseridos automaticamente no sistema de pesquisa;

Requisitos Éticos e Legais

Quem possui os direitos autorais dos dados?

- Eu;

Informe a licença que será aplicada aos dados.

- CC-BY-NC

Como serão gerenciadas as questões éticas e legais?

No Termo de Consentimento Livre Esclarecido serão incluídas informações sobre as possibilidades de compartilhamento e publicação dos dados e uso por outros pesquisadores de forma anônima, assim como os dados que possam identificar individualmente os participantes não serão divulgados. Todos os dados que possam revelar a identidade dos participantes serão anonimizados e terão identificadores exclusivos atribuídos a eles. Um arquivo separado com dados identificadores ou semi-identificadores serão mantidos separados em segurança em computador com acesso restrito e por meio de senha e backup em nuvem institucional e não serão compartilhados ou publicados.

Você obteve consentimento para o compartilhamento, a preservação e a reutilização dos dados por terceiros?

- Sim

Compartilhamento de Dados

Você planeja compartilhar os dados gerados/coletados na pesquisa?

- Sim

Quando você irá compartilhar os dados?

- Após a conclusão da pesquisa;

Onde você vai compartilhar os dados?

- Repositório de dados
- Solicitações diretas
- Solicitações diretas

Os dados ficarão embargados?

- Não

Responsabilidades e Recursos

Descreva as funções e responsabilidades de todas as atividades, por exemplo, captura de dados, produção de metadados, documentação, armazenamento e backup, arquivamento e compartilhamento de dados. Os responsáveis devem ser nomeados sempre que possível.

Todos os processos serão realizados pelo responsável pela pesquisa.

APÊNDICE C – PROTOCOLO DE COLETA DE DADOS DAS BIBLIOTECAS ESTRANGEIRAS

- 1 – Investigar se a biblioteca possui política para a gestão de dados científicos.
- 2 – Investigar se a biblioteca possui divulgação de informações sobre gestão de dados científicos por meio de seus sites, redes sociais ou e-mail institucional.
- 3 – Investigar se a biblioteca possui palestras já realizadas sobre gestão de dados científicos.
- 4 – Investigar se a biblioteca possui profissionais capacitados para a oferta de SPGDC.
- 5 – Investigar se a biblioteca possui serviços ou produtos sobre os Princípios FAIR.
- 6 – Investigar se a biblioteca possui serviços ou produtos sobre o ciclo de vida dos dados
- 7 – Investigar se a biblioteca possui serviços ou produtos sobre o Plano de Gestão de Dados
- 8 – Investigar se a biblioteca possui serviços ou produtos sobre dados pessoais e sensíveis.
- 9 – Investigar se a biblioteca possui serviços ou produtos para orientar os pesquisadores a encontrar dados úteis para suas pesquisas
- 10 – Investigar se a universidade possui repositório institucional de dados científicos.
- 11 – Investigar se a biblioteca possui serviço ou produto para indicar repositórios de dados científicos aos pesquisadores da universidade.

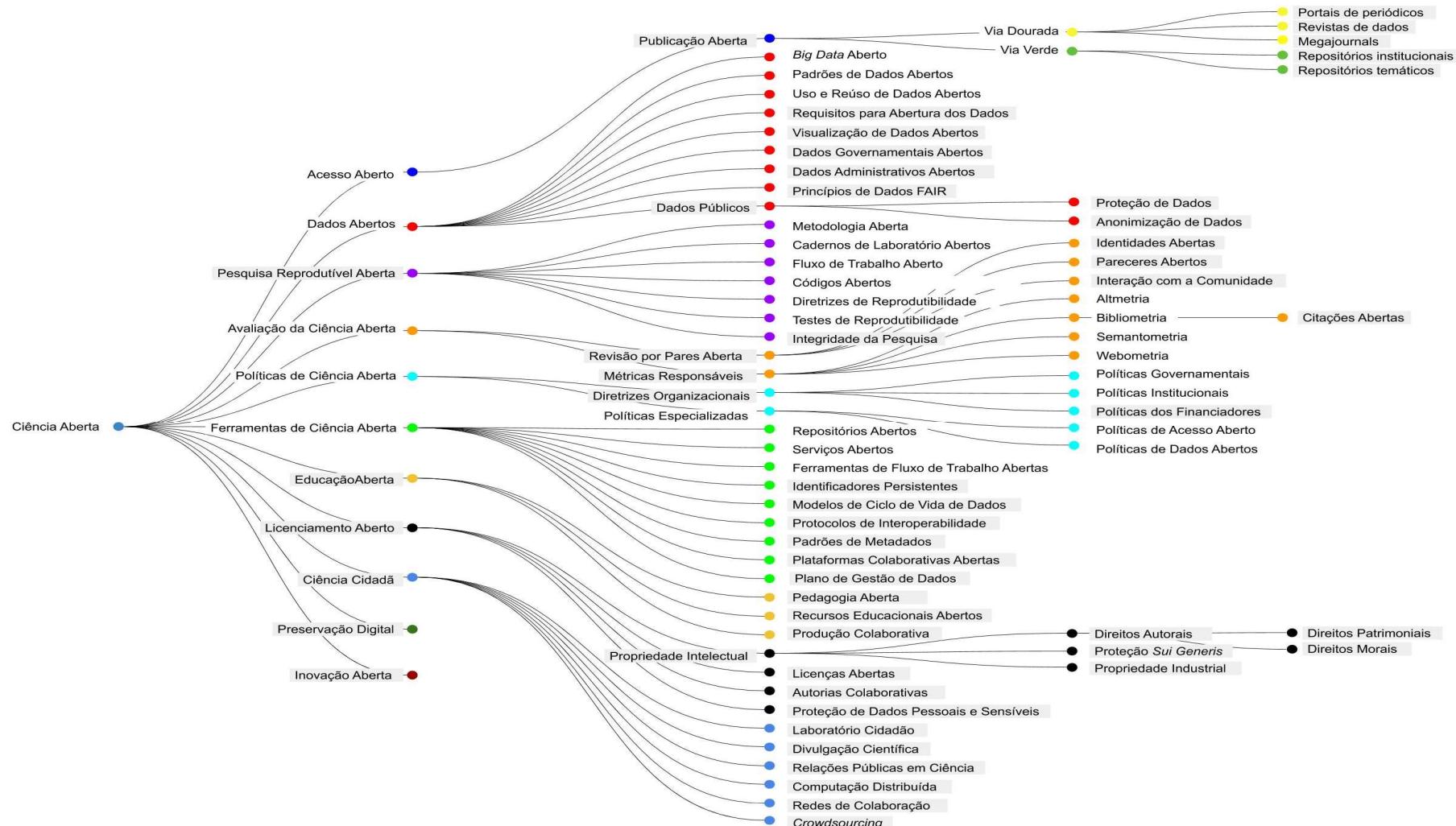
APÊNDICE D - CONJUNTO DE SPGDC SEGUNDO A LITERATURA CIENTÍFICA

Autores	SPGDC
Fearon <i>et. al.</i> (2013)	<ul style="list-style-type: none"> • Política • Orientação para construção do PGD • Software para construção do PGD • Modelos de PGD • Orientação sobre direitos autorais • Orientação para localização de dados para uso em pesquisas científicas • Orientação da gestão de dados durante à pesquisa • Curadoria de dados (seleção, tratamento, compartilhamento, preservação, segurança, arquivamento). • Orientação para escolha dos metadados • Orientação para documentar os dados • Orientação para encontrar repositórios de dados científicos específicos (temáticos ou interdisciplinares) • Capacitação: treinamentos, cursos, palestras • Tutoriais • Repositório de dados científicos • Arquivamento de dados
Jones; Pryor; White, (2013)	<ul style="list-style-type: none"> • Planejamento de GDC • Política • Disponibilização de informações sobre GDC em site • Consultorias • <i>Helpdesk</i> • Treinamentos para o corpo docente • Treinamentos para o profissional bibliotecário • Suporte na construção do PGD • Software para construção do PGD • Suporte a GDC durante à pesquisa • Armazenamento de dados • Armazenamento de dados em nuvem • Seleção e entrega de dados • Ferramentas para criação de metadados • Orientação para depósito em repositórios externos • Repositório de dados científicos
	<ul style="list-style-type: none"> • Consulta sobre Plano de Gestão de Dados (PGD)

Tenopir (2016)	<ul style="list-style-type: none"> • Consultoria sobre padrões de metadados • Divulgação sobre GDC • Criação de guias • Auxiliar os pesquisadores na GDC durante o projeto de pesquisa • Suporte para encontrar e citar conjuntos de dados • Promoção de eventos para discutir GDC • Treinamento para o corpo técnico da biblioteca • Participação no desenvolvimento de políticas para GDC • Suporte na preparação dos dados e conjuntos de dados para depósito em repositório • Criação ou transformação de metadados para dados e conjuntos de dados
Silva (2016)	<ul style="list-style-type: none"> • Oficina para PGD • Orientação para escolha dos metadados • Orientação para escolha de repositório mais adequado para o depósito dos conjuntos de dados • Orientação para escolha de soluções de armazenamento de dados • Preservação de dados • Orientação sobre o ciclo de vida dos dados • Orientação para documentar os dados • Orientação para escolha da licença de uso • Suporte para encontrar dados e conjunto de dados • Re却tório de dados científicos
Tang; Hu (2019)	<ul style="list-style-type: none"> • Planejamento de GDC • Descoberta e acesso a dados • Organização de dados • Curadoria de dados • Documentação de dados • Metadados • Compartilhamento e disseminação de dados • Preservação de dados • Visualização de dados • Gerenciador de citações de dados • Software para gerenciamento de dados durante a pesquisa • Motor de busca de dados • Software de processamento de dados

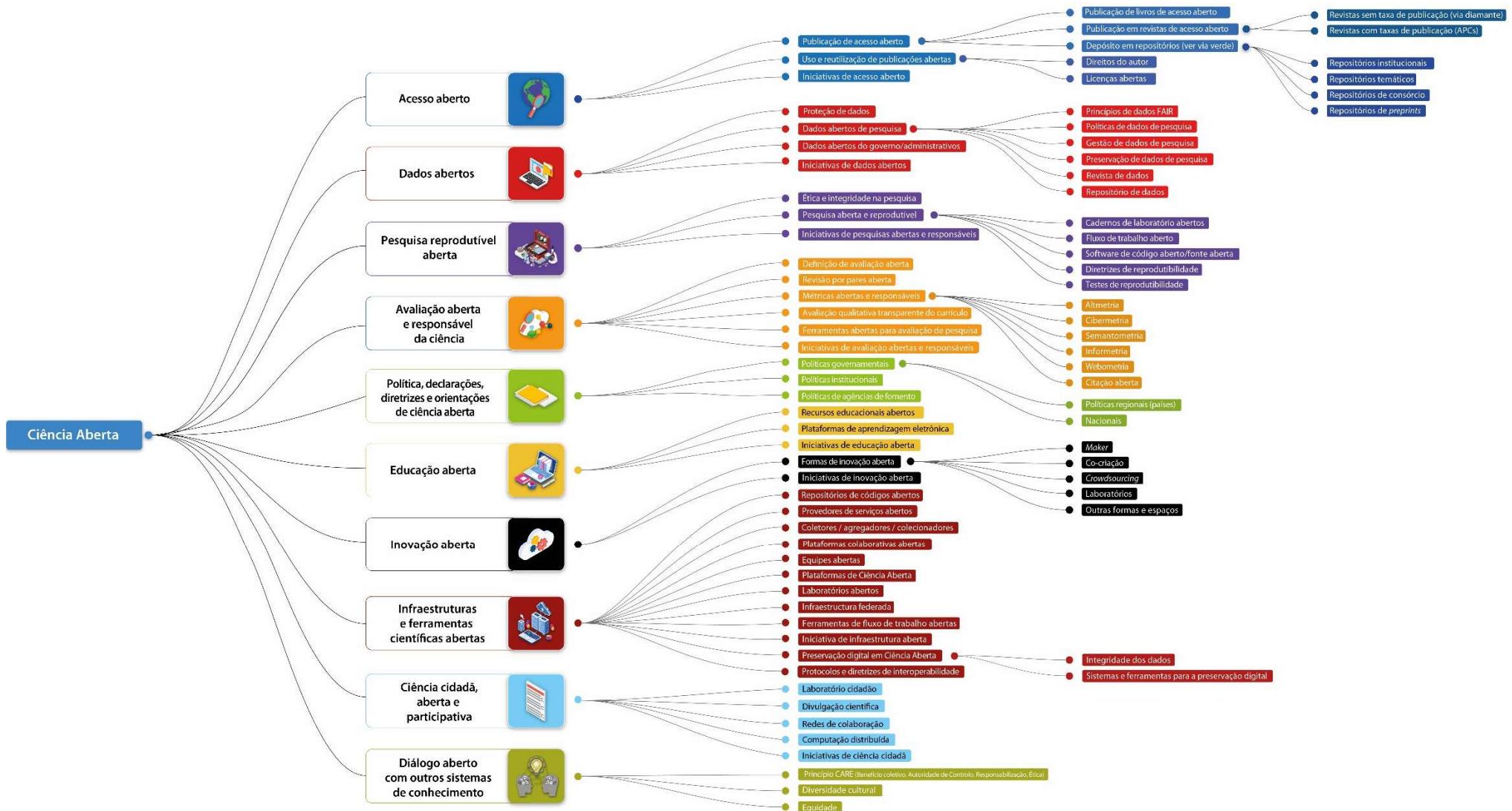
Semeler; Pinto, (2020)	<ul style="list-style-type: none"> • Repositório de dados científicos • Curadoria de dados • Análise, extração e entrega de dados • Técnicas de arquivamento • Técnicas de preservação de dados • Gerenciamento de repositório de dados • Suporte na documentação de dados • Oficinas • Armazenamento seguro de dados • Suporte na construção do PGD • Treinamento em organização e documentação de dados • Treinamento para capacitar bibliotecários • Visualização de dados
Huang; Cox; Sbaffi (2020)	<ul style="list-style-type: none"> • Orientação sobre PGD • Treinamentos • Informações sobre GDC em site • Orientação sobre citação de dados • Orientação sobre publicação de dados • Orientação sobre armazenamento de dados • Orientação sobre análise de dados • Orientação sobre mineração de dados • Orientação sobre visualização de dados • Curadoria de dados • Preservação • Repositório de dados científicos • Criação de metadados • Catalogação de dados • Seleção de dados • Preparação de dados • Verificação da qualidade dos dados • Suporte na GDC durante a pesquisa científica

ANEXO A – TAXONOMIA DA CIÊNCIA ABERTA POR PESQUISADORES BRASILEIROS



Fonte: Silveira *et al.* (2021)

ANEXO B – TAXONOMIA DA CIÊNCIA ABERTA 2023



Fonte: Silveira *et al.* (2023).