



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

NORMACI CORREIA DOS SANTOS SENA

**GESTÃO DE DADOS DE PESQUISA CIENTÍFICA NA
PERSPECTIVA DA CIÊNCIA ABERTA:
um estudo de caso na Universidade Federal da Bahia**

Salvador
2023

NORMACI CORREIA DOS SANTOS SENA

**GESTÃO DE DADOS DE PESQUISA CIENTÍFICA NA
PERSPECTIVA DA CIÊNCIA ABERTA:
um estudo de caso na Universidade Federal da Bahia**

Tese apresentada ao Programa de Pós-Graduação em Ciência da Informação, Instituto de Ciências da Informação, Universidade Federal da Bahia como requisito para obtenção do grau em Doutora em Ciência da Informação.

Área de concentração: Políticas e Tecnologias da Informação

Orientador: Prof. Dr. Raymundo das Neves Machado

Salvador

2023

S474g Sena, Normaci Correia dos Santos.

Gestão de dados de pesquisa científica: um estudo de caso na
Universidade Federal da Bahia / Normaci Correia dos Santos Sena. -
Salvador, 2023.
200 p.; il.color

Orientador: Prof. Dr. Raymundo das Neves Machado
Tese (doutorado) - Universidade Federal da Bahia. Instituto de
Ciência da Informação, Salvador, 2023.

1. Gestão de dados de pesquisa. 2. Ciência Aberta. 3. Universidade Federal da Bahia - dados de pesquisa. I. Machado, Raymundo da Neves.
- II. Universidade Federal da Bahia. Instituto de Ciência da Informação.
- III. Título.

CDD (23 ed.) 020

NORMACI CORREIA DOS SANTOS SENA

Gestão de dados de pesquisa científica na perspectiva da Ciência Aberta: um estudo caso na Universidade Federal da Bahia

Tese apresentada ao Curso de Doutorado 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 Doutor em Ciência da Informação.

Aprovada em: 29 de maio de 2023.

Banca Examinadora

Documento assinado digitalmente
 RAYMUNDO DAS NEVES MACHADO
Data: 30/05/2023 15:59:27-0300
Verifique em <https://validar.it.gov.br>

Prof. Dr. Raymundo das Neves Machado - Orientador – UFBA

Documento assinado digitalmente
 FABIANO COUTO CORRÊA DA SILVA
Data: 28/05/2023 21:59:41-0300
Verifique em <https://validar.it.gov.br>

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

Documento assinado digitalmente
 FLÁVIA GOULART MOTA GARCIA ROSA
Data: 27/05/2023 15:37:59-0300
Verifique em <https://validar.it.gov.br>

Profº. Dra. Flávia Goulart Mota Garcia Rosa - Membro Externo Titular – UFBA

Documento assinado digitalmente
 GILLIAN LEANDRO DE QUEIROGA LIMA
Data: 01/06/2023 13:44:35-0300
Verifique em <https://validar.it.gov.br>

Prof. Dr. Gillian Leandro de Queiroga Lima - Membro Externo Titular – UFBA



Profº. Dra. Nidia Maria Lienert Lubisco - Membro Interno Titular – UFBA

A

Alexsandro, meu esposo, por ter caminhado comigo;
Eduardo e Júlia, filhos amados, por ter sido meus maiores
incentivadores;
Minha família, pelos ensinamentos e acolhimento nos momentos
delicados.

AGRADECIMENTOS

A continuidade do estudo não foi minha primeira opção, na verdade, nem cogitava fazer o mestrado, imagine o doutorado, mas pela necessidade no trabalho e por incentivo de pessoas, aqui estou finalizando mais uma etapa vitoriosa em minha vida. As dificuldades poderiam ter sido maiores, mas não senti tanto porque, modéstia parte, sempre fui uma pessoa disciplinada, organizada e ansiosa para exercer minhas atividades com muita eficiência; mas de que adianta ser assim, se não tiver as contribuições de pessoas singulares em sua vida? E são a elas que registro meus agradecimentos...

À Deus, por ser meu guia, por me conceder saúde física, mental e espiritual para lidar com as pressões muitas vezes causadas por mim mesmo, e principalmente por me dar sabedoria e colocar pessoas iluminadas nesse percurso para caminharmos juntos.

À minha família, esposo Alexsandro, por toda paciência, amor e incentivo; e aos meus filhos, Eduardo e Júlia, por serem meus grandes incentivadores, orgulho e amores da minha vida. Por eles, respirei e segui.

Ao meu orientador e amigo, professor Raymundo das Neves Machado, pelo apoio, dedicação, paciência e parceria em todas as fases, principalmente, por tornar esse caminho mais leve.

À minha banca examinadora, professores doutores Fabiano Couto, Nídia Lubisco, Alegria Benchimol, Flávia Goulart, José Cláudio e Gillian Queiroga pelo aceite e pelas contribuições necessárias para o fortalecimento e enriquecimento da ciência.

À Editora da Universidade Federal da Bahia (EDUFBA), a porta de entrada para me tornar a profissional que sou hoje.

Aos meus colegas e amigos da Biblioteca Professor Jorge Portugal da Câmara Municipal de Salvador pela paciência e incentivo.

À Casa do Estudante Lucas Silva (CELS), a primeira porta para tornar esse título possível.

Aos colegas e amigos da pós-graduação, em especial a Weslayne, minha parceira de perrengues acadêmicos. Obrigada pela escuta e pelo acolhimento nos momentos diversos.

A todos os pesquisadores que fizeram parte desta pesquisa e a todos que de alguma forma contribuíram para este resultado: os colegas do doutorado, colegas de trabalho e outros.

SENA, Normaci Correia dos Santos. **Gestão de dados de pesquisa científica**: um estudo de caso na Universidade Federal da Bahia. 2023. Orientador: Raymundo das Neves Machado. 200 f. il. Tese (Doutorado em Ciência da Informação) - Instituto de Ciência da Informação, Universidade Federal da Bahia, Salvador, 2023.

RESUMO

O movimento da Ciência Aberta traz à tona discussões relevantes acerca da ciência colaborativa do século XXI. Nesse universo coparticipativo, os dados de pesquisas têm sido objetos de estudo em virtude da necessidade social de se obter dados de confiança para o avanço científico de forma aberta e gratuita. Além disso, diversas agências de fomento reconhecem a importância dos dados de pesquisa e, consequentemente, começaram a exigir a sua adequada gestão. Nesse contexto, enquanto grandes produtoras e consumidoras de dados de pesquisas, as universidades são protagonistas na produção e na disseminação do conhecimento, fortalecendo as pesquisas e o processo de inovação, e ratificando o compromisso social e democrático das instituições públicas. Ao contextualizar a literatura sobre dados de pesquisa, percebe-se uma lacuna nas universidades brasileiras na estruturação de políticas que direcionem os pesquisadores à gestão dos dados científicos, já que o movimento da Ciência Aberta aponta, em suas vertentes, um tratamento adequado dos dados de pesquisa de forma a viabilizar o processo de armazenamento, organização, busca, recuperação e difusão dos dados coletados. E isso impacta na adesão (ou não) do pesquisador na abertura e no compartilhamento de sua produção científica, dentre outros motivos. Assim, este estudo tem por objetivo geral: conhecer a percepção dos professores pesquisadores no que tange à gestão de dados e às ações da Universidade Federal da Bahia (UFBA) desenvolvidas para a implantação de políticas de gestão de dados de pesquisas no contexto da Ciência Aberta. Trata-se de um estudo de caso único integrado, uma vez que, no Brasil, os estudos a respeito de plano de gestão de dados de pesquisa científica ainda se encontram em estágios embrionários. A saber da importância histórica que é a Universidade Federal da Bahia, esta foi escolhida para este estudo. Para o percurso metodológico, optou-se como instrumento de coleta de dados o questionário on-line seguindo o modelo da Escala Likert. Por se tratar de uma pesquisa qualitativa, a análise dos dados foi realizada sob a técnica da Lógica Paraconsistente, com a finalidade de tratar e equilibrar os graus de concordância e discordância entre as proposições, evitando falhas interpretativas. E, em conjunto, fez uso do coeficiente de correlação de Spearman. A pesquisa identificou algumas lacunas na estrutura de políticas para a gestão de dados nas universidades brasileiras, apesar do reconhecimento crescente da importância dos dados de pesquisa no avanço da ciência. Além disso, revelou que a UFBA ainda está distante das práticas de ciência aberta observadas mundialmente. Os pesquisadores indicaram a necessidade de políticas que apoiem a gestão de dados de pesquisa e enfatizaram a falta de diretrizes e apoio institucional nesse processo. Portanto, foi possível alcançar um melhor entendimento sobre a percepção e práticas dos pesquisadores brasileiros em relação à gestão de dados de pesquisa, bem como um olhar crítico sobre a falta de políticas efetivas para apoiar a Ciência Aberta no contexto das universidades brasileiras. Além disso, ressalta a urgência de aprimorar e implementar um plano de gestão de dados de pesquisa na UFBA, o que pode ter implicações em outras universidades e instituições de pesquisa no Brasil.

Palavras-chave: Gestão de dados de pesquisa científica; Universidade Federal da Bahia; dados de pesquisa; política de gestão de dados; ciência aberta.

SENA, Normaci Correia dos Santos. **Scientific research data management**: a case study at the Federal University of Bahia. 2023. Advisor: Raymundo das Neves Machado. 200 f. il. Thesis (Doctorate in Information Science) - Institute of Information Science, Federal University of Bahia, Salvador, 2023.

ABSTRACT

The Open Science movement brings up relevant discussions about collaborative science in the 21st century. In this collaborative universe, research data has been the object of study due to the social need to obtain reliable data for scientific advancement in an open and freeway. In addition, several development agencies recognize the importance of research data, consequently, they began to demand adequate management of these data. This highlights the need to ensure means and mechanisms for preserving and reusing data through policies that regulate the opening of data from public resources. By contextualizing the literature on research data, a gap is perceived in Brazilian universities as well as in most funding agencies in the structuring of policies that direct researchers to the management of scientific data, since the Open Science movement points to an appropriated treatment of research data to enable the process of storage, organization, search, recovery, and dissemination of collected data. And this impacts on the researcher's adherence (or not) in opening and sharing their scientific production, among other reasons. Thus, this study has the general objective: to know the perception of research professors regarding data management and UFBA actions developed for the implementation of a research data management policies in the context of Open Science. This is a single integrated case study, since, in Brazil, studies regarding data management plans for scientific research are still in their embryonic stages. Knowing the historical importance of the Federal University of Bahia, it was chosen for this study. For the methodological route, the online questionnaire was chosen as an instrument for data collection, following the Likert Scale model. As this is a qualitative research, data analysis was carried out using the Paraconsistent Logic technique, with the aim of treating and balancing the degrees of agreement and disagreement between propositions, avoiding misinterpretations. And, together, the Spearman correlation coefficient was used. The research identified some gaps in the policy framework for data management in Brazilian universities, despite the growing recognition of the importance of research data in the advancement of science. In addition, he revealed that UFBA is still far from the open science practices observed worldwide. The research identified some gaps in the policy framework for data management in Brazilian universities, despite the growing recognition of the importance of research data in the advancement of science. In addition, he revealed that UFBA is still far from the open science practices observed worldwide. The researchers indicated the need for policies that support the management of research data and emphasized the lack of guidelines and institutional support in this process. Therefore, it was possible to achieve a better understanding of the perception and practices of Brazilian researchers in relation to the management of research data, as well as a critical look at the lack of effective policies to support Open Science in the context of Brazilian universities. In addition, it highlights the urgency of improving and implementing a research data management plan at UFBA, which may have implications for other universities and research institutions in Brazil.

Keywords: Scientific research data management; Federal University of Bahia; search data; data management policies; open science.

LISTA DE FIGURAS

Figura 1 - Facetas da Ciência Aberta.....	32
Figura 2 - Taxonomia da Ciência Aberta.....	37
Figura 3 - Uso do <i>preprint</i> durante a pandemia da Covid-19.....	44
Figura 4 - Modelo de ciclo de vida dos dados da <i>Digital Curation Centre</i> (DCC).....	74
Figura 5 - Taxonomia de gestão de dados de pesquisa.....	79
Figura 6 - Princípios FAIR.....	89
Figura 7 - Atividades da GO FAIR.....	91
Figura 8 - Protocolo para a coleta de dados.....	109
Figura 9 - Delineamento da pesquisa e estratégia metodológica.....	110

LISTA DE QUADROS

Quadro 1 - Benefícios da Ciência Aberta na sociedade contemporânea.....	29
Quadro 2 - Definições de Ciência Aberta.....	30
Quadro 3 - Ranking das Universidades brasileiras no período de 2017-2020 no quesito publicação de acesso aberto.....	53
Quadro 4 - Fundações de Amparo à Pesquisa no Brasil.....	81
Quadro 5 - Requisitos exigidos pelo CoretrustSeal em 2020-2022 para avaliação de repositório de dados.....	86
Quadro 6 - Cursos de Pós-Graduação da UFBA com respectivo conceito Capes no quadriênio 2017-2021.....	100
Quadro 7 - Etapas para a coleta de dados.....	111
Quadro 8 - Relação das proposições conforme a categoria pretendida e sua relação com o referencial teórico.....	114
Quadro 9 - Interpretação de valores de GCp.....	121
Quadro 10 - Diretrizes estratégicas do PDI da UFBA , 2018-2022.....	126
Quadro 11 – Ciência Aberta.....	131
Quadro 12 – Matriz de associação de Spearman entre as proposições da categoria “Ciência Aberta”.....	132
Quadro 13 – Gestão de dados de pesquisa.....	132
Quadro 14 – Matriz de associação de Spearman entre as proposições da categoria “Gestão de dados de pesquisa”.....	133
Quadro 15 – Atitudes e valores em relação a abertura de dados.....	134
Quadro 16 – Matriz de associação de Spearman entre as proposições da categoria “Atitudes e valores em relação a abertura de dados”.....	135
Quadro 17 – Motivação para a abertura.....	136
Quadro 18 – Matriz de associação de Spearman entre as proposições da categoria “Motivação para a abertura”.....	137
Quadro 19 – Necessidades dos pesquisadores.....	138

Quadro 20 – Matriz de associação de Spearman entre as proposições da categoria “Necessidades dos pesquisadores”.....	138
Quadro 21 – Agência de fomento.....	139
Quadro 22 – Matriz de associação de Spearman entre as proposições da categoria “Agência de fomento”.....	140
Quadro 23 – Direito autoral/Proteção de dados.....	141
Quadro 24 – Matriz de associação de Spearman entre as proposições da categoria “Direito autoral/Proteção de dados”.....	141
Quadro 25 – Necessidades de profissional capacitado e/ou Biblioteca.....	142
Quadro 26 – Matriz de associação de Spearman entre as proposições da categoria “Necessidade de profissional capacitado e/ou Biblioteca”.....	143
Quadro 27 – Razões para não compartilhar dados de pesquisa.....	144
Quadro 28 – Matriz de associação de Spearman entre as proposições da categoria “Razões para não compartilhar dados de pesquisa”.....	145
Quadro 29 – Repositório de dados.....	147
Quadro 30 – Matriz de associação de Spearman entre as proposições da categoria “Repositório de dados”.....	148
Quadro 31 - Síntese dos graus de concordância com base na lógica paraconsistente.....	149

LISTA DE TABELA

Tabela 1 – População da pesquisa.....	103
Tabela 2 - Relação de respondentes por área de conhecimento.....	129
Tabela 3 - Relação de professores por faixa etária.....	129
Tabela 4 - Relação de professores por tempo de trabalho na Instituição/UFBA.....	130

LISTA DE ABREVIATURAS E SIGLAS

AADP	Acesso Aberto a Dados de Pesquisa
ANDS	Australian National Data Service
BDEP	Banco de Dados de Exploração e Produção
BDTD	Biblioteca Digital Brasileira de Teses e Dissertações
BOAI	Iniciativa de Acesso Aberto de Budapeste (BOAI - Budapest Open Access Initiative, em inglês)
Brapci	Base de Dados Referenciais de Artigos de Periódicos em Ciência da Informação
CA	Ciência Aberta
Capes	Coordenação Nacional de Aperfeiçoamento de Pessoal de Nível Superior
CC	Creative Commons
CIGA	Comitê Interministerial Governo Aberto
CNPq	Conselho Nacional de Desenvolvimento Científico e Tecnológico
CoNCiênciA	Consórcio Nacional para Ciência Aberta
CT&I	Ciência, Tecnologia e Inovação
CVDP	Ciclo de vida dos dados de pesquisa
DANS	Data Archiving and Networked Services
DCC	Digital Curation Centre
DDI	Data Documentation Initiative
DOI	Digital Object Identifiers
DSA	Data Seal of Approval
EaD	Educação a Distância
Embrapa	Empresa Brasileira de Pesquisa Agropecuária
EOSC	European Open Science Cloud
FAIR	findable, accessible, interoperable e reusable
FAP	Fundações Estaduais de Amparo à Pesquisa
Fapesp	Fundação de Amparo à Pesquisa no Estado de São Paulo
Finep	Financiadora de Estudos e Projetos
Fiocruz	Fundação Oswaldo Cruz
Foster	Facilitate Open Science Training for European Research
FURG	Universidade Federal do Rio Grande

GLOBE	Global Collaboration Engine
IBICT	Instituto Brasileiro de Informação em Ciência e Tecnologia
IODP	International Ocean Discovery Program
JDDCP	Joint Declaration Data Citation Principles
LAI	Lei de Acesso à Informação
LDA	Lei de Direitos Autorais
LP	Lógica Paraconsistente
MCTI	Ministério da Ciência, Tecnologia e Inovação
M/D	Mestrado/Doutorado
MIT	Massachusetts Institute of Technology
MP	Mestrado Profissional
NIH	National Institutes of Health
OGP	Open Government Partnership - Parceria para Governo Aberto
OECD	Organização para Cooperação e o Desenvolvimento Econômico
PDA	Plano de Dados Abertos
PDI	Plano de Desenvolvimento Institucional
PELD	Repositório de Dados do Programa de Pesquisas Ecológicas de Longa Duração
PGD	Plano de Gestão de Dados
PI	Propriedade Intelectual
PPBIO	Repositório de Dados do Programa de Pesquisa de Biodiversidade da Amazônia Ocidental
RDP Brasil	Rede de Dados de Pesquisa Brasileira
RNP	Rede Nacional de Ensino e Pesquisa
SciELO	Scientific Electronic Library Online
SEC	Secretaria da Educação
STJU	Shangai Jiao Tong University
TCLE	Termo de Consentimento Livre e Esclarecido
TICs	Tecnologias de Informação e Comunicação
UE	União Europeia
UFBA	Universidade Federal da Bahia
UFRGS	Universidade Federal do Rio Grande do Sul

UNEB Universidade Estadual da Bahia
Unesp Universidade Estadual Paulista Júlio de Mesquita
USP Universidade de São Paulo

SUMÁRIO

1 INTRODUÇÃO.....	17
2 CIÊNCIA ABERTA E SUAS INTERFACES.....	26
2.1 COMUNICAÇÃO CIENTÍFICA.....	38
2.2 CIÊNCIA ABERTA NO MUNDO E NO BRASIL.....	47
2.2.1 Propriedade intelectual e autoria colaborativa	55
3 GESTÃO DE DADOS DE PESQUISA.....	60
3.1 DADOS.....	61
3.2 DADOS DE PESQUISA.....	65
3.2.1 Ciclo de vida dos dados de pesquisa científica.....	72
3.2.2 Gestão e plano de dados de pesquisa.....	77
3.2.3 Princípios FAIR e repositórios de dados de pesquisa.....	85
4 PERCURSO METODOLÓGICO.....	95
4.1 ABORDAGEM METODOLÓGICA.....	95
4.2 UNIVERSO, AMOSTRA E POPULAÇÃO-ALVO DA PESQUISA.....	98
4.3 PROCEDIMENTOS.....	104
4.3.1 Estudo de caso.....	104
4.3.2 Elaboração do instrumento de coleta de dados.....	110
4.3.2.1 Questionário : teste piloto.....	118
4.3.3 Procedimentos de coleta de dados.....	123
5 ANÁLISE E DISCUSSÃO DOS DADOS.....	124
5.1 ANÁLISE DAS POLÍTICAS PARA GESTÃO DE DADOS DE PESQUISA NA UFBA.....	124
5.2 ANÁLISE DO QUESTIONÁRIO.....	128
6 CONSIDERAÇÕES FINAIS.....	151
REFERÊNCIAS.....	155
APÊNDICE A - Apresentação e Termo de Consentimento Livre e Esclarecido.....	168
APÊNDICE B - Questionário aplicado aos professores de pós-graduação.....	169
APÊNDICE C - Projeto piloto de Plano de Gestão de Dados de Pesquisa para a UFBA.....	175
APÊNDICE D - Plano de Gestão de Dados.....	186
ANEXO A - Plano de dados abertos da UFBA.....	191

1 INTRODUÇÃO

A Sociedade da Informação e/ou do Conhecimento foi um grande marco no processo evolutivo das atividades intrínsecas ao ser humano, principalmente nas transformações decorrentes das Tecnologias de Informação e Comunicação (TICs) e das atividades científicas. Embora haja muitas discussões sobre o conceito e interpretação desse tipo de sociedade, no contexto contemporâneo, a informação e o conhecimento fazem surgir revoluções científicas capazes de fornecer contributos teóricos, metodológicos e filosóficos a respeito da informação e da informação tecnológica. Enquanto a Sociedade da Informação refere-se à ideia da “inovação tecnológica”, Sociedades do Conhecimento engloba uma dimensão de transformação social, cultural, econômica, política e institucional, assim como uma perspectiva mais pluralista e de desenvolvimento, ou seja, expressa a complexidade e o dinamismo das mudanças que estão ocorrendo (BURCH, 2005). Nessa perspectiva, advém dessa Era uma nova ordem científica, que enfatiza a responsabilidade sociocientífica e prioriza a disponibilização, em acesso aberto, de dados e informações, isto é, mesmo que timidamente, o acesso à informação promove, de forma consciente, a semente da liberdade.

Se, por um lado, a revolução tecnocientífica possibilita o livre acesso à produção e à comunicação científica para a sociedade, por outro, rompe com os sistemas tradicionais de comunicação científica ao tempo que privilegia a natureza colaborativa; consequentemente, causa insegurança e resistência dos pesquisadores aos novos *modus operandi* de comunicar a pesquisa em decorrência da cultura comercial do século XX que está enraizada.

A partir da segunda metade do século XX – período marcado pela dicotomia entre sigilo e transparência e a expansão do conhecimento científico decorrente das inovações ocasionadas pela Segunda Guerra Mundial (1939-1945) –, originou-se a Ciência Aberta (CA) ou *e-Science*, também denominada por Costa (2006) de filosofia aberta. Ela ampliou novas formas de pensar, produzir e comunicar cientificamente, permitindo não apenas a abertura dos dados, mas a discussão de todo o processo de sua produção, desde os *softwares* utilizados até as etapas de avaliação, possibilitando a discussão de métricas alternativas à ciência. Nesse ínterim, viu-se a disseminação dos dados de pesquisa como uma das condições para a avaliação dos resultados científicos, pois a exposição e tudo que foi utilizado em um trabalho de pesquisa pode evidenciar erros e incertezas, que o pesquisador não apontou em suas publicações (ALBAGLI, 2015).

Ao falar de Ciência Aberta, no Brasil, é necessário destacar a lei que marca o início de uma Era, quando o acesso à informação pública se tornou obrigatória na administração pública – a Lei de Acesso à Informação nº 12.527, de 18 de novembro de 2011. Essa lei traz iniciativas de atividades de movimento de abertura de dados, permitindo o estímulo ao desenvolvimento de novos produtos e serviços no âmbito público ou privado. Resultante dessa lei, o Plano de Dados Abertos (PDA) prevê a abertura de dados (BRASIL, 2016; PARCERIA...2013, 2016a, 2016b, 2018); isto é, a disponibilização de dados em formato bruto, os quais as pessoas possam livremente acessar, manipular e reutilizar, ampliando, com isso, a participação social.

Paralelamente, os modelos tradicionais de comunicação da pesquisa, no cerne da Ciência, começaram a ser amplamente debatidos; ou seja, “[...] as iniciativas mais expressivas nos estágios iniciais do movimento pela Ciência Aberta dirigiram-se centralmente para o acesso livre a publicações científicas” (ALBAGLI, 2015, p. 13). Para a autora, o interesse pela propriedade intelectual aumentou a pressão nos ambientes intelectuais e universitários. Em suas palavras:

[...] o movimento pela ciência aberta se insere nesse quadro de tensão entre, por um lado, novas formas de produção colaborativa, interativa e compartilhada da informação, do conhecimento, da cultura. E, por outro, mecanismos de captura e privatização desse conhecimento que é coletiva e socialmente produzido (ALBAGLI, 2015, p. 13).

Tal fato permitiu que os mecanismos de apropriação privada se ampliassem como um movimento de compartilhar o conhecimento entre a “[...] comunidade científica, a sociedade e as empresas, possibilitando desta forma ampliar o reconhecimento e o impacto social e econômico da ciência” (ALBAGLI, 2015, p. 14). Esse movimento se baseia em práticas de colaboração e produção do conhecimento, em consonância com as ferramentas digitais que disponibilizam, não apenas artigos, mas dados científicos, notas de laboratório, avaliações de pares, projetos, entre outros.

Amparada nos fundamentos sociológicos – produto de colaboração social – e econômico – ao considerar que as produções científicas geradas com recursos públicos são bens públicos –, a Ciência Aberta é um movimento de transformações cultural e social. Ou seja, integra ferramentas com foco na colaboração aberta, acesso aberto a publicações, recursos educacionais, *software* de código aberto, ciência cidadã, acesso e reúso de dados abertos de pesquisas, avaliação e revisão aberta.

Seguindo os rumos das tendências internacionais e reconhecendo a relevância do acesso aberto na produção intelectual e dados científicos, muitas universidades brasileiras têm promovido e se inserido nesse movimento. Estudos recentes mostram que tanto o governo quanto as agências de fomento brasileiras e internacionais já iniciaram recomendações – ou exigências – para que os projetos financiados englobem um plano de gestão de dados de pesquisa, visando ao compartilhamento em acesso aberto (BERTIN *et al.*, 2019b; BRINEY; GOBEN; ZILINSKI, 2015; COUNCIL ON LIBRARY AND INFORMATION RESOURCES, 2013; CORRÊA, 2019).

Uma boa gestão informacional assegura que os conjuntos de dados produzidos pelos pesquisadores sejam compartilhados e reaproveitados em outras pesquisas, com transparência e reproduzibilidade. Para uma boa prática de pesquisa, é essencial que os dados sejam coordenados e compartilhados, contribuindo no avanço científico e tecnológico, ratificando o papel social da Ciência.

Com base na literatura, a disponibilização dos dados levantados e produzidos pelas pesquisas são requisitos para uma ciência, de fato, aberta. Dados que deixam de ser meros subprodutos das atividades de pesquisa e passam a ser foco de grande magnitude para todo o meio científico (BIAZON; MARIN, 2016; SAYÃO; SALES, 2016). A comunicação científica envolve todo processo da pesquisa, isto é, desde a ideia inicial até a aceitação dos resultados pela comunidade.

Os efeitos do acesso aberto e da Ciência Aberta têm ampliado as alternativas propostas por editores científicos a fim de que objetos digitais de pesquisas estejam acessíveis. No que se refere aos dados de pesquisa, reforça-se a necessidade do acesso a eles a fim de potencializar o investimento público. Nesse viés, a disponibilização dos dados de pesquisa, ou seja, a possibilidade de acesso, seja público ou privado, independe de um processo formal e isso difere da publicação dos dados, já que para publicar é necessário redigir conforme convenções usuais de cada área do conhecimento, e visa possibilitar a outros pesquisadores a replicação dos experimentos (TORINO; ROA-MARTÍNEZ; VIDOTTI, 2020). Assim, o conhecimento científico é um patrimônio da humanidade, e, em vista disso, os dados de pesquisas precisam estar disponíveis livremente para que as pessoas possam usá-los e distribuí-los.

Sabe-se que, ao longo do tempo, no Brasil, o movimento de institucionalização da ciência mudou para as universidades e institutos públicos, tornando-se os principais ambientes

de produção do conhecimento, e o poder Estatal tornou-se seu principal financiador. À medida que evoluiu, o movimento pela Ciência Aberta transformou-se e inseriu novos elementos, transpassando a questão do acesso livre a publicações científicas, e incorporando práticas de acesso aberto, que viabilizam abertura e compartilhamento dos dados.

Parte-se do pressuposto de que a Ciência Aberta é um indicativo para verificar mudanças nas práticas de pesquisa, e que o uso intensivo de dados de pesquisa científica, bem como o de infraestrutura altamente tecnológica e colaborativa, envolve discussões extensivas no âmbito das políticas de ciência. Percebe-se, com isso, o reconhecimento do potencial informacional dos dados de pesquisa pelos pesquisadores, instituições acadêmicas e agências de fomento.

Em um contexto de pesquisa e inovação e considerando que as universidades, enquanto instituições de pesquisa, precisam adaptar-se, preparar-se e atender a essa nova demanda informacional, e que os cientistas também têm como função publicizar as suas atividades de pesquisa, uma das características desse movimento de ação colaborativa de acesso livre é a definição de políticas, incluindo a gestão de dados e tecnologias digitais, tendo como centro as pessoas.

Sabe-se que as universidades são grandes protagonistas da construção, disponibilização do conhecimento e consumidoras de dados de pesquisas. Tais dados fortalecem as pesquisas e o processo de inovação, e ratificam o compromisso social e democrático das instituições públicas. Ao mesmo tempo, têm-se dois embates: de um lado, “[...] os esforços pelo amplo compartilhamento dos resultados da pesquisa científica”; na outra ponta, “[...] interesses favoráveis a pesquisas de cunho proprietário e que são movidas principalmente pela lucratividade [...]” (ALBAGLI; APPEL; MACIEL, 2013, p. 8).

Tradicionalmente, a relação entre poder e ciência sempre foi conflitante. Isso porque o pesquisador cientista precisa negociar com o poder (governo, empresas), a fim de obter recursos para a profissionalização de áreas; no entanto, gera os riscos de manipulações da ciência. Essa lógica de custos chegou também às universidades em meados dos anos 1980, e no mesmo século em que se iniciou o movimento de acesso aberto na comunicação científica, conforme afirmou Marcos Cueto¹ em declaração publicada na revista *História, Ciência, Saúde – Manguinhos* (HCSM), em 2019. Não obstante, só em 1990 é que esse conceito se expandiu, e as instituições internacionais e nacionais ampliaram as diretrizes para a abertura

¹ Editor-científico da revista *História, Ciências, Saúde – Manguinhos*.

dos dados nas pesquisas, enfatizando e reconhecendo o poder informacional introduzido por meio da Ciência Aberta (MONTEIRO; LUCAS, 2019).

As tecnologias digitais, assim como a evolução mundial para criação de políticas públicas de informação², fortaleceram a necessidade de transparência e o direito à informação de caráter público, principalmente se essas informações são produzidas mediante custos de dinheiro público. Nas universidades não é diferente. Fazer Ciência Aberta implica permitir que outros contribuam e colaborem com a pesquisa. O fato é que, embora esse movimento tenha ganhado força na ciência, ainda há rejeição por parte dos pesquisadores. Porém, o fenômeno da comunicação dos dados de pesquisa é um elemento capaz de melhorar não apenas o fluxo da comunicação, mas também dos processos da ciência, contribuindo para o debate teórico da área. Pesquisa publicada em 26 de fevereiro de 2021 confirma isso; o acesso aberto superou, em todo mundo, pela primeira vez, os modos tradicionais de publicação, fortalecendo a inserção ao movimento da Ciência Aberta por parte da comunidade científica, especialmente em países como China, Alemanha e Reino Unido; enquanto que o Brasil não é citado na referida pesquisa (HOOK, 2021).

Alinhado a isso, em 2022, agências financeiras como, por exemplo, o Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) tem fomentado a criação de um Consórcio Nacional para Ciência Aberta (CoNCiência), uma forma de estimular e fortalecer o compartilhamento de dados de pesquisa. A iniciativa pretende convocar instituições que objetivam ter Digital Object Identifiers (DOI) relacionados a seus conjuntos de dados (*datasets*), para isso, foi realizado um acordo com a *datacite* (organização internacional que gera o DOI). Além disto visa criar seu próprio repositório de dados (*lattes data*) a fim de que seus beneficiários depositem ali os seus dados de pesquisa relativos a projetos financiados pelo Conselho³.

Sendo os dados de pesquisa um dos elementos essenciais que compõem a Ciência Aberta, a viabilidade deles visa facilitar descobertas e avanços científicos. Esses avanços exigem das universidades adoção de políticas de gestão de dados a fim de obter benefícios que vão além da durabilidade de um projeto de pesquisa; isto porque a sociedade científica se

² Trata-se de um “conjunto de premissas, decisões e ações - produzidas pelo Estado e inseridas nas agendas governamentais em nome do interesse social - que contemplam os diversos aspectos (administrativo, legal, científico, cultural, tecnológico, etc.) relativos à produção, uso e preservação da informação de natureza pública e privada.” (JARDIM; SILVA; NHARRELUGA, 2009, p. 9).

³ Mais informações ver:

<https://www.gov.br/cnpq/pt-br/acesso-a-informacao/acoes-e-programas/partners/cooperacao-internacional/conselho-nacional-para-ciencia-aberta>

encontra atualmente num profundo e célere processo de transformação e o compartilhamento dos dados devem ocorrer de forma organizada de maneira que facilite o reuso dos dados.

Considerando que as universidades possuem grande papel para a sociedade no acesso ao conhecimento, esta pesquisa é relevante porque evidenciará e reforçará o impacto na sociedade e a responsabilidade sociocientífica na gestão e abertura dos dados de pesquisas desenvolvidas pela Universidade Federal da Bahia (UFBA). A escolha pela Instituição se deu devido ao reconhecimento pelo seu pioneirismo nos estudos interdisciplinares na Bahia e também pela pesquisadora residir no Estado.

No contexto da Ciência da Informação, contribuirá para discussões e práticas na promoção do movimento da Ciência Aberta, no âmbito das Universidades – nesse caso, a UFBA –, tendo como reflexo os novos *modus* da comunicação científica para a sociedade, uma vez que já há exigências do plano de gestão de dados por parte das financiadoras de pesquisas no Brasil. A Fundação de Amparo às Pesquisas no Estado de São Paulo (Fapesp), por exemplo, exige o plano na submissão de projetos temáticos. No entanto, pela velocidade natural, espera-se que outras fundações de pesquisas estaduais também iniciem as exigências. Ademais, as demandas de pesquisas científicas são cada vez mais necessárias e crescentes por gestão de dados ativos e heterogêneos. Portanto, é imprescindível que haja bases de conhecimentos científicos mais eficientes nas representações do conhecimento, de maneira que os fenômenos, as dimensões e as dinâmicas possam ser melhor inseridos, ou seja, uma gestão dos dados mais eficiente.

A ideia de trabalhar esta temática surgiu como um desdobramento do resultado da dissertação, cujo título foi “Profissional da informação no contexto de dados abertos nos legislativos da cidade de Salvador, Ba: uma análise a partir da lógica paraconsistente”, onde foi abordado a atuação do bibliotecário na perspectiva de uma das facetas da Ciência Aberta. Contudo, senti necessidade de estudar o acesso à informação no universo acadêmico e aprofundar sobre dados científicos; além disso, as mudanças mundiais para participação efetiva na abertura dos dados de pesquisas possibilitaram um cenário oportuno para a realização deste estudo. Diante do exposto, percebe-se que se deve ter um olhar mais refinado para a relevância e conhecimento do processo de gestão de dados de pesquisas nas universidades e, principalmente, a compreensão dos pesquisadores nesse movimento. Isso porque a literatura evidencia uma resistência por parte do pesquisador em compartilhar os seus dados, mesmo havendo políticas mandatórias. Observa-se, também, pouco envolvimento

das instituições acadêmicas brasileiras na criação e organização de gestão dos dados de pesquisas científicas (VEIGA, 2017; VEIGA *et al.*, 2019).

Assim, este trabalho tem como fio condutor o seguinte questionamento: como a Universidade Federal da Bahia tem desenvolvido as ações, no contexto da Ciência Aberta, para a implantação de políticas de gestão de dados de pesquisa, e quais as percepções e os desafios dos professores nesse processo?

Nesse sentido, a pesquisa tem por objetivo geral: conhecer a percepção dos professores pesquisadores no que tange à gestão de dados e às ações da UFBA desenvolvidas para a implantação de políticas de gestão de dados de pesquisas no contexto da Ciência Aberta. Os objetivos específicos são:

- OE1: verificar na literatura os principais conceitos da Ciência Aberta, dados de pesquisa, plano de gestão de dados e FAIR⁴;
- OE2: mapear as ações desenvolvidas para a implantação de uma política de gestão de dados de pesquisa na UFBA;
- OE3: identificar a percepção dos pesquisadores a respeito dos fatores que podem influenciar a abertura e/ou compartilhamento de dados de pesquisa científica;
- OE4: descrever os principais desafios e as soluções inerentes à construção de uma política estruturada para a gestão de dados de pesquisa científica na UFBA.

A partir das fundamentações teóricas, este estudo se enquadra no chamado quarto paradigma científico, no qual se baseia a Ciência Aberta – inovadora e fundamentada nos usos e reússos cada vez mais intensos dos dados de pesquisa –, que aos poucos está sendo inserida no Brasil. Para Thomas Kuhn (1922-1996), os paradigmas, além de nortear a atividade de uma comunidade científica, são essenciais para delinear a fase em que uma área de pesquisa se encontra, contribuindo para explicar as revoluções científicas. Essas mudanças de paradigmas são motivadas, muitas vezes, pelos trabalhos de cientistas que com suas práticas e teorias possibilitam o avanço na maneira de fazer e comunicar a ciência.

A mudança de paradigmas, segundo Kuhn (2011) se dá por práticas reforçadas pelos dogmas e teorias antigas que se transformaram em decorrência da evolução histórica, científica, econômica, política, cultural e social na sociedade. No contexto da ciência, essas

⁴ FAIR (*Findable, Accessible, Interoperable e Reusable*).

transformações acontecem com as inserções de novos conceitos e definições de teorias e modelos já existentes. Nesse contexto, o que Thomas Kuhn propõe não se trata apenas de uma mudança de paradigmas, mas uma evolução científica, pois muda o modo como a ciência caracteriza no mundo.

Assim, diferentemente das ciências tradicionais que eram pautadas na experimentação (1º paradigma) – período em que os filósofos e estudiosos explicavam as suas observações através de leis naturais e não causas sobrenaturais; na ciência teórica (2º paradigma) – alguns cientistas como Isaac Newton fizeram previsões para os novos fenômenos e verificações de novas hipóteses através das realizações de experimentos, ocasionando novas teorias; e na ciência computacional (3º paradigma) – fase em que trata a simulação de fenômenos complexos para a exploração científica; a Ciência Aberta (4º paradigma) utiliza dos dados primários adjacentes de diferentes disciplinas, utilizados como insumos na investigação científica (OLIVEIRA, 2016).

Tradicionalmente, a ciência esteve pautada na observação, contudo a ciência contemporânea está voltada para a prática colaborativa e engloba ferramentas e recursos tecnológicos e digitais colaborativos, convergindo para o uso dos dados científicos, por isso denomina-se como o quarto paradigma.

De fato, a ciência e a tecnologia, somadas à comunicação científica e à Ciência Aberta, estão embasadas no argumento de que os dados podem ter vários usos e, portanto, proporcionam mais visibilidade dos resultados desse processo. Consequentemente, maiores chances de serem obtidos mais conhecimentos científicos. Ademais, considera-se a ideia de que os resultados de pesquisas oriundos do financiamento público são bens públicos tangíveis; logo, devem estar disponíveis sem custo para reúso (DALRYMPLE, 2003).

Assim, pretende-se realizar um estudo de caso na Universidade Federal da Bahia, apoiado na premissa de que o plano de gestão de dados facilita a sua compreensão e a reutilização futura. Contudo, é imprescindível conhecer a percepção daqueles que criam e coletam esses dados: os pesquisadores. Além disso, é necessário entender as políticas já implantadas, que tangenciam a abertura e gestão desses dados de pesquisa. Cabe às universidades e agências de fomento à pesquisa ou centros nacionais a responsabilidade quanto ao desenvolvimento dessas políticas.

Considera-se, com isso, que o elo semântico entre a produção científica e dados de pesquisa, bem como a disponibilização dos dados brutos e/ou processados, são importantes

para o fortalecimento da ciência e da democracia, pois abre possibilidades de colaboração científica. Dessa forma, justifica-se a realização deste estudo, uma vez que possibilita contribuições para a Ciência da Informação, no que se refere à comunicação e disseminação da informação, corroborando os anseios da sociedade civil e acadêmica.

Dessarte, esta tese ilustra as atividades que foram desenvolvidas no decurso do doutoramento. O primeiro capítulo refere-se a esta apresentação; no segundo, discorrer sobre o movimento da Ciência Aberta e suas interfaces; no terceiro, são abordados assuntos sobre a gestão de dados de pesquisa e discussões referentes ao tema; no quarto, apresenta-se o percurso metodológico; o quinto, são apresentados os resultados e discussão dos dados, seguido das referências, dos apêndices e dos anexos.

2 CIÊNCIA ABERTA E SUAS INTERFACES

A origem da ciência moderna foi caracterizada por mudanças sociais, políticas e econômicas. Uma junção de interesses e oportunidades circunstanciais venceram o período medieval marcado pelo segredo, culminando num conjunto de convenções sociais e artifícios institucionais que moldaram a República da Ciência – termo criado por Michael Polanyi, em 1962 –, em que denominava essa República de sociedade dos exploradores; isto é, empenham-se na busca de um futuro desconhecido, compromissados com sua própria curiosidade e a busca de sua satisfação intelectual (MERTON, 1973; SCHWARTZMAN, 2008). O nascimento desse período foi marcado pela rápida abertura e a ampla disseminação das novas descobertas e invenções; além disso, a República da Ciência estimulou não apenas a racionalidade individual, mas a ciência como um todo.

Muitas características da instituição República da Ciência foram oriundas da República das Letras (*Republic of Letters*) – movimento direcionado à criação, disseminação e aplicação do que se denomina filosofia experimental –, ou seja, a ciência atual. Tal fato explica, segundo Guimarães (2014, p. 3), o início da Ciência Aberta, embora a compreensão generalizada desse termo seja recente:

[...] formada por um grupo restrito de cidadãos letRADOS, é definida por pesquisadores a partir de diferentes ângulos: uma comunidade de sábios; o conteúdo das ideias que a comunidade abraça; o meio de disseminar as ideias; a instituição que determina padrões de persuasão (pertinência da prova e reproduzibilidade do experimento); conjunto de atitudes visando à colaboração e a abertura, entre outros.

A história mostra que o homem da ciência era o homem das letras, que, por meio da comunicação escrita, fazia circular os inventos, os experimentos que vivenciava em suas viagens. Esse intelectual tinha como norma a publicidade do conhecimento; contudo,

[...] a norma básica da comunidade era que todo conhecimento deveria ser colocado na esfera pública tão logo fosse gerado. O criador manteria o direito de propriedade sobre o conhecimento, mas esse direito não incluía aquele de excluir terceiros de usá-lo (GUIMARÃES, 2014, p. 3).

Conforme exposto, a abertura dos dados informacionais era a maneira pela qual o conhecimento tinha mais confiabilidade, pois foi uma forma de tornar o conhecimento mais confiável, na proporção em que os críticos e os céritos questionavam (MOKYR, 2008). Por outro lado, Guimarães (2014), ao citar Paulo David (2004), trouxe à tona a lógica econômica da ciência aberta; ou seja, por meio de incentivos, os talentosos mantinham a prática de produção de conhecimento:

[...] a abertura não se prestava somente ao reconhecimento e prestígio dos inventores por meio de seus pares; mais precisamente, era uma certificação que permitia que os mesmos fossem reconhecidos pelos patronos e recebessem patrocínio (GUIMARÃES, 2014, p. 141).

O movimento da Ciência Aberta abarca diferentes pilares que definem o seu conceito, não se limitando apenas a facilitar o livre acesso a publicações científicas e a dados de pesquisa. A CA engloba o desenvolvimento de aspectos amplos e complexos, os quais cooperam para uma ciência colaborativa que responda aos anseios da sociedade moderna. Em outras palavras, trata-se de uma reforma de cultura com a finalidade de tornar a ciência mais transparente, acessível e reproduzível.

Em sua própria essência, entende-se Ciência Aberta como conhecimento transparente – na verdade, aberto –, e que transcende os campos científicos por meio das tecnologias digitais, alcançando a sociedade. É um movimento que estimula a transparência da pesquisa científica, desde a geração da investigação até o uso de *software* aberto (SILVA; SILVEIRA, 2019). Para compreender a estrutura dessa ciência, é importante destacar alguns termos que a literatura aborda, ao se referir às atividades de pesquisas colaborativas e intensivas em dados e informação. Todos eles são sinônimos de CA e encontram-se bem presentes em outros países, mas que apresentam múltiplas interpretações, por exemplo: *Big Data*, *e-Research*, *e-Science*, *e-infra-estructure* e *cyber infra-estructure*, *e-Humanities*.

Para Piorun (2013), *e-Science*, *e-Humanities* e *e-social sciences* referem-se a uma disciplina específica ou a um conjunto de disciplinas, enquanto os termos *e-Research* ou *cyber scholarship* são mais frequentemente utilizados dentro do contexto de todas as disciplinas acadêmicas; ou seja, representam o conceito mais amplo. Com o avanço na infraestrutura cibernética, pesquisadores estão colaborando com disciplinas e instituições para abordar questões de pesquisa e hipóteses que foram, anteriormente, difíceis de serem examinadas em decorrência do escopo, magnitude e complexidade dos tópicos estudados.

No que se refere às denominações apresentadas, Albagli, Appel e Maciel (2013) não concordam que *Big Data* seja sinônimo de CA, até porque se observarmos a figura na página 35, percebe-se que é subordinado a uma das vertentes da CA (*Dados Abertos*), portanto, há uma distinção entre eles. Para eles, *e-Science* é um termo usado na comunidade científica que faz uso intensivo de dados, cuja gênese é a pesquisa científica colaborativa. Em contrapartida, *Big Data* refere-se à permuta e análise de grandes volumes de dados, a fim de extrair informações de dados não estruturados; melhor dizendo, a *Big Data* é consequência da *e-Science* e, portanto, a gestão desses dados de pesquisas precisa ser de responsabilidade não

apenas dos pesquisadores, mas das universidades e das bibliotecas.

Na perspectiva de Piorun (2013), a pesquisa baseada em dados é referida nas ciências físicas, biológicas e médicas, em larga escala, distributivas, computacionalmente intensa, como *e-Science*. O Centro Nacional de Ciência *e-Science*, do Reino Unido, comprova essa assertiva ao definir *e-Science*:

Ciência em larga escala, que será cada vez mais realizada por meio de colaborações globais distribuídas pela Internet. Normalmente, um aspecto de tais colaborações são empreendimentos científicos que exigem acesso a coleções *big data*, a recursos de computação em grande escala e a visualização de alta performance, tendo os cientistas como usuários individuais (PIORUM, 2013, p. 3, tradução direta).

Portanto, a *e-Science* não se trata apenas de uma nova disciplina científica, mas também de um conjunto de ferramentas e tecnologias relevantes para auxiliar a Ciência em rede e a sua infraestrutura, com a finalidade de capacitar os pesquisadores a desenvolverem suas pesquisas de modo mais eficaz e colaborativo. Assim, por uma conveniência conceitual, o termo utilizado neste trabalho é Ciência Aberta, entendendo que se trata de um meio interposto por um ambiente colaborativo e aberto.

Para a Comissão Europeia, a ciência sempre esteve aberta, e o que acontece atualmente é um amplo reconhecimento de acessibilidade à produção científica. Consequentemente, apresenta outros desafios que precisam ser enfrentados – infraestrutura, direitos de propriedade intelectual, mineração de conteúdo e métricas alternativas –, e também a colaboração interinstitucional, interdisciplinar e internacional entre todos os atores em pesquisa e inovação. Trata-se de um conceito em evolução, mas que mobiliza pontos de vista distintos e muitas vezes antagônicos (ALBAGLI, 2015; EUROPEAN COMMISSION, 2020).

Sem consenso sobre sua extensão e seu significado, Albagli (2015) mostra, então, que o formato atual do movimento pela Ciência Aberta reflete novas formas de pensar e de exercer a científicidade, ressoando nos compromissos, normas e arcabouços institucionais que interferem diretamente na prática científica e nas suas relações com a sociedade. Se, por um lado, a CA aumenta a visibilidade, o acesso e a velocidade da produção e circulação do conhecimento científico, por outro, evidencia a base social da ciência, conferindo maior porosidade na sua relação e interlocução com outros tipos de saberes e agentes cognitivos. Para Albagli (2015), portanto, é preciso uma concepção mais democrática que abarque outros atores e espaços de conhecimento, e não apenas uma perspectiva limitada à abertura ao campo científico e a um novo tipo de produtivismo em ciência.

Nessa perspectiva, entende-se que a maneira como a ciência atual ainda é realizada não atende a valores básicos da comunicação científica, como por exemplo: o registro da descoberta, velocidade da publicação, solidez nos dados, replicabilidade e reproduzibilidade das pesquisas, e também sua ampla distribuição. Em virtude disso, as temáticas em torno da Ciência Aberta preconizam o acesso aberto às publicações ao tempo que possibilita inúmeros benefícios para os participantes da comunidade científica, pesquisadores, financiadores de pesquisa bem como para a comunidade em geral, como se ver abaixo:

Quadro 1 - Benefícios da Ciência Aberta na sociedade contemporânea

Pesquisadores	Comunidade científica	Financiadores de pesquisa	Público em geral
visibilidade das pesquisas; reconhecimento pessoal; aumenta a reputação; credibilidade; possibilidade de encontrar novas parcerias de pesquisa; aumenta a probabilidade de obter financiamentos; mais acessibilidade; e, aumenta o número de citações e atenção por parte da mídia.	diminui os custos operacionais de coleta, tabulação e manipulação de dados de pesquisas; aumenta a colaboração entre agentes que coletam dados e agentes que utilizam dados; aumenta a possibilidade de se replicar pesquisas publicadas; e, auxilia a elaboração de materiais didáticos de pesquisa	ajuda a entender e a controlar a alocação dos recursos, públicos e privados, destinados à pesquisa e avaliar o seu retorno econômico; diminui a probabilidade de gastos em pesquisas ou experimentos duplicados ou muito semelhantes.	ajuda os agentes decisores a formular melhores e mais eficientes políticas públicas e a melhorar o debate político; aumenta o direito democrático de acesso a dados e conhecimento por parte de todos os participantes da sociedade, o que permite maior engajamento do público na atividade científica.

Fonte: Martins (2020).

Contudo, mesmo com os benefícios em questão, há necessidade de ampliar o debate a respeito da consolidação desse movimento, tendo em vista que a resistência técnica e, principalmente, política ainda é alta. Isso porque, segundo Martins (2020), “precisa-se de um sistema que, simultaneamente, aumente a transparência das pesquisas sem comprometer a liberdade individual dos cientistas e sem incentivar comportamentos antiéticos por parte de não-cientistas.” Ou seja, todas essas questões demonstram fragilidade e falta de conhecimento sobre a CA por parte dos atores envolvidos nesse movimento e frisa a necessidade de ampliar a discussão a respeito.

Nesse contexto, a crescente demanda para tornar a pesquisa aberta modifica e envolve novos elementos que compõem práticas ligadas às tecnologias digitais colaborativas, bem

como instrumentos da propriedade intelectual. Em resposta a essa evolução, várias tentativas foram realizadas a fim de definir Ciência Aberta que, de certa maneira, se complementam e abrangem novas diretrizes nesse movimento. Como exemplo, estão as seguintes definições: do projeto Facilitate Open Science Training for European Research (FOSTER), com viés mais socioeconômico; das Fundações Open Society, evento realizado para marcar o décimo aniversário da Iniciativa de Acesso Aberto de Budapeste (BOAI); e outras definições com viés mais econômico e democrático (ALBAGLI; CLINIO; RAYCHTOCK, 2014), conforme exposto no Quadro 2. Cabe ressaltar que a Iniciativa de Acesso Aberto de Budapeste, em 2002, foi responsável por estabelecer o acesso aberto pela primeira vez como uma prática de publicação internacional, quando incluiu-se também a reutilização de todos os tipos de produtos de pesquisa e não apenas publicação ou dados.

Quadro 2 - Definições de Ciência Aberta

CONCEITO DE CIÊNCIA ABERTA	INSTITUIÇÃO/AUTOR	TERMO DE DESTAQUE
“ Acesso aberto à literatura científica revisada por pares significa a disponibilidade livre na Internet, permitindo a qualquer usuário ler, fazer <i>download</i> , copiar, distribuir, imprimir, pesquisar ou referenciar o texto integral desses artigos, recolhê-los para indexação, introduzi-los como dados em <i>software</i> , ou usá-los para outro qualquer fim legal, sem barreiras financeiras, legais ou técnicas que não sejam inseparáveis ao próprio acesso a uma conexão à Internet”.	BUDAPESTE OPEN ACCESS INITIATIVE (2002)	Acesso aberto Uso Gratuidade Sem restrições
“A Ciência Aberta é a ideia de que conhecimentos científicos de todos os tipos deveriam ser compartilhados abertamente tão logo quanto possível no processo de descoberta.”	Nielson (2011)	Conhecimento Compartilhamento Descoberta
“[...] significa muitas coisas, mas principalmente que o conhecimento científico deve ser livre para as pessoas usarem, reutilizarem e distribuírem sem restrições legais, tecnológicas ou sociais”.	Albagli, Clinio e Raychtock (2014)	Livre Uso Distribuição Sem restrições Reúso
“A prática da ciência de tal forma que		

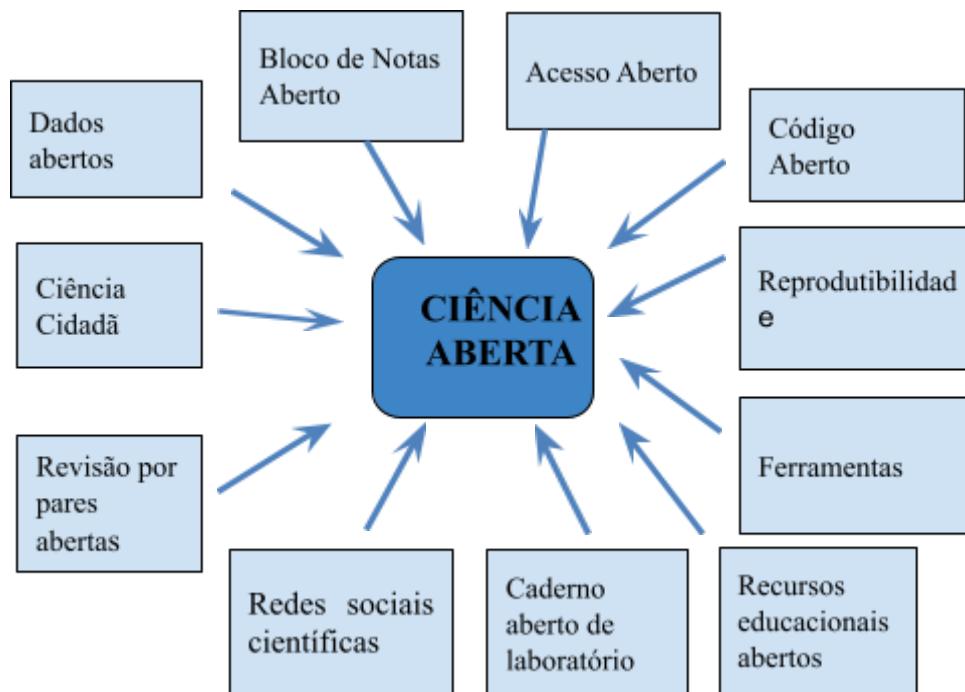
<p>outros podem colaborar e contribuir, na qual os dados de pesquisa, as notas de laboratório e outros processos de pesquisa estão disponíveis livremente, em condições que permitem a reutilização, redistribuição e reprodução da pesquisa e dos dados e métodos subjacentes”.</p> <p>“[...] o conhecimento científico é produto da colaboração social e por isso sua propriedade pertence à comunidade. Pelo ponto de vista econômico, resultados científicos gerados a partir de pesquisas públicas são bens públicos que qualquer um poderia ter acesso sem nenhum custo”.</p>	Foster (2018a)	Colaboração Contribuição Livre Uso Reúso Conhecimento
<p>“[...] nova abordagem colaborativa, transparente e acessível à pesquisa, o que implica uma mudança estrutural na forma de conceber pesquisas e disseminar seus resultados. De forma simples, é fazer com que os resultados da pesquisa financiada com fundos públicos sejam acessíveis em formato digital, não só para a comunidade científica que os produz, mas para a sociedade em geral que o financia, aumentando a reprodutibilidade da ciência e a reutilização dos resultados.”</p>	Mendéz (2018)	Colaborativa Transparente Acessível Disseminação Reprodutibilidade Uso Reúso Gratuidade
<p>“[...] a Ciência Aberta traz consigo uma série de mudanças que vêm revolucionando o modo de se fazer a pesquisa científica”.</p>	Sá e Campos (2017)	Revolução
<p>“[...] para fazer as principais saídas de resultados de pesquisa financiados publicamente – publicações e dados de pesquisa – acessíveis publicamente em formato digital sem restrição ou mínima”.</p>	OECD (2015)	Acessível Público Gratuito

Fonte: Elaboração própria.

Conforme exposto, o Quadro 2 mostra a definição de Ciência Aberta em um espaço temporal de 2015 a 2018, envolvendo diferentes atividades além do acesso e reúso dos dados. Segundo se vê, trata-se de um movimento dentro de outros movimentos. Por isso, é popularmente conhecido como um termo guarda-chuva, pois abarca diversos pilares e múltiplas dimensões, com o intuito de remover as barreiras para compartilhar as produções, recursos, métodos ou ferramentas em qualquer etapa do processo de pesquisa (Figura 1). Essas atividades, fortalecidas pela Internet, separam a ciência em duas etapas: antes da rede e depois da rede. Essa nova Era (depois da rede), em que há o rápido compartilhamento e envio das informações, abre espaço às publicações eletrônicas e dá uma nova roupagem à

comunicação científica.

Figura 1 - Facetas da Ciência Aberta



Fonte: Adaptado de *Foster Open Science*

Acesso Aberto: A definição mais usada originou-se em 2001, quando foi criada a Declaração da Iniciativa de Budapeste pelo Acesso Aberto. Nesse documento, acesso aberto à literatura científica significa disponibilidade livre na Internet, aberto e gratuito, sem quaisquer barreiras (custos, registro/login e outros), ou restrições ao acesso associado aos direitos autorais (patrimoniais) e licenciamento, ou seja, com poucas ou nenhuma limitações à reutilização⁵.

O acesso aberto consiste em duas vias que asseguram o acesso: a publicação de revistas de acesso aberto (também conhecida como via dourada) e o autoarquivamento/depósito das publicações em repositórios de acesso aberto (conhecida como via verde). A primeira abrange os periódicos científicos eletrônicos, cujo acesso aberto a seus conteúdos é garantido pelos próprios editores; isto é, não usa os direitos de autor para restringir o acesso e uso do material que publicam. A segunda possibilita o autoarquivamento, que poderá ser realizado pelos próprios autores de artigos científicos já publicados ou aceitos para publicação; dessa maneira, existe permissão dos editores que os aceitaram para que os possam disponibilizar em um servidor de arquivo aberto.

⁵ Ver: <https://www.budapestopenaccessinitiative.org/boai-10-recommendations>

Dados abertos: para a *Open Knowledge International*, dados são abertos quando qualquer pessoa pode livremente acessá-los, compartilhá-los, modificá-los e utilizá-los para qualquer finalidade. São dados brutos de pesquisa científica, de forma organizada e sistemática, registrada em bibliotecas de dados abertos. Portanto, as informações públicas disponíveis em formato aberto devem ser para uso livre, sem restrições de licenças ou qualquer outro mecanismo de controle⁶.

Ciência cidadã (*Citizen science*): É uma das práticas da CA bastante desafiadora, pois trata-se não apenas da valorização, mas também da participação direta de não cientistas e não especialistas no fazer ciência, como “leigos” e “amadores” da ciência cidadã (ALBAGLI; CLINIO; RAYCHTOCK, 2014). Os autores destacam a relevância da autoria colaborativa em uma pesquisa. Para eles, quanto mais colaboradores, mais informações e conhecimentos são gerados, propiciando novos usos desses conhecimentos. “É o envolvimento do público não acadêmico no processo de pesquisa científica – seja pesquisa orientada para a comunidade ou a nível global”⁷. Trata-se, portanto, de maior participação dos cidadãos de forma voluntária na realização dos trabalhos científicos. O *Galaxy Zoo*⁸ é um exemplo desse tipo de cooperação. Por mais de uma década, voluntários têm participado desse projeto com a finalidade de classificar as galáxias. Outro exemplo é o *E-Bird*⁹, gerenciado pelo Cornell Lab of Ornithology, o maior projeto de ciência cidadã, relacionado à biodiversidade do mundo. Por meio dele, os observadores de aves incluem quando, onde e como as aves foram vistas e ouvidas, e assim o aplicativo coleta dados de qualquer lugar do mundo. É um projeto global, do qual o Brasil participa.

A respeito do funcionamento dessa prática, em 2013, foi produzido pelo Socientize Consortium¹⁰, um relatório que descreve e realiza a compilação de experiências de ciência cidadã na Europa. Nesse documento, além de encontrar distintas definições da ciência cidadã, indicam-se suas condições de desenvolvimento e as barreiras existentes para seu fomento em direção às políticas de inovação científica e tecnológica. Na prática, na ciência cidadã, o público em geral participa das atividades de pesquisa científica, contribuindo ativamente com seu esforço intelectual e conhecimento local, além de suas ferramentas e recursos, resultando na abertura transdisciplinar, e conectado em redes, com aprimoração e integração da ciência,

⁶ Para mais informações ver: <https://okfn.org/opendata/>

⁷ A respeito, ver: <https://www.citizenscience.org/>

⁸ Para mais informações ver: <http://zoo1.galaxyzoo.org/>

⁹ Para mais informações ver: <https://ebird.org/home>

¹⁰ Trata-se de um Consórcio que coordena uma consulta pública em curso e debate a respeito do potencial papel da ciência cidadã na Europa.

política e sociedade (SOCIENTIZE CONSORTIUM, 2013; PARRA, 2015).

Revisão por pares aberta: para Foster (2018b), “[...] a revisão por pares aberta visa, portanto, trazer maior transparência e participação aos processos de revisão por pares formais e informais”. Esse processo está ligado a dois componentes fundamentais na publicação científica: a comunicação e a avaliação de ciência. Tradicionalmente, essa revisão é demorada, anônima, seletiva, fechada, e os revisores não recebem recompensa financeira para a realização dessa atividade.

Na revisão por pares aberta, o revisor sabe quem é o autor e vice-versa; há menores hipóteses para preconceitos ou ações inapropriadas; os revisores podem obter crédito pelo seu trabalho; além disso, é possível animosidade dos autores e repercussões em oportunidades de emprego, promoção e financiamento. Assim, ao contrário do sistema de revisão por pares tradicional – que nem sempre consegue garantir qualidade da investigação comunicada –, a revisão por pares aberta oferece maior exposição e maior número de potenciais revisores.

Código aberto: conforme mencionado anteriormente, a Ciência Aberta não se limita à forma das publicações abertas; é preciso que o conteúdo também seja aberto. Esse é um dos pilares da Ciência Experimental, que nada mais é do que a sua reproduzibilidade, regras de procedimentos que permitem a obtenção de resultados claros e precisos. E ser reproduzível significa que os dados e ferramentas aplicados nos experimentos, simulações e análises sejam também disponibilizados de forma aberta e livre. Para isso, necessita-se de Código Aberto – ou seja, *Software Livre*.

A Fiocruz, portanto, entende código aberto ou *Software* de Código Aberto (OSS) como “[...] um tipo de *software* cujo código-fonte é liberado sob uma licença na qual o detentor dos direitos autorais concede aos usuários os direitos para estudar, alterar e distribuir o *software* para qualquer pessoa e para qualquer finalidade”¹¹. Podendo ser desenvolvido de maneira pública colaborativa.

Caderno aberto em laboratórios: trata-se de um registro primário da pesquisa científica, em que os pesquisadores registram suas experiências, hipóteses, análises e interpretações dos resultados. Além disso, serve como ferramenta de proteção à propriedade intelectual resultante da pesquisa. Sobre isso, Antonio Lafuente (2014), ao propor sobre o caderno aberto de laboratório cidadão, estima a abertura do conhecimento de forma que não se restrinja ao acesso e à distribuição aberta, on-line e gratuita apenas, mas também a novas

¹¹ Curso Ciência Aberta - Série 1 - Aula 4 - Workflows abertos (fiocruz.br).

maneiras de produção de conhecimento.

Para Lafuente (2014, on-line), há muita ostentação de projetos a que se intitulam abertos e/ou participativos, mas não os são: “[...] muitos casos em que a ação de abertura e/ou participação são apenas rodovias que levam ao fortalecimento de novos regimes de propriedade e à privatização do público”. Ainda, para o referido autor, “[...] a abordagem comum não se destina a explorar os caminhos do conhecimento aberto [...] a participação, o novo imperativo de nossas sociedades democráticas, tem como aspiração melhorar a funcionalidade do sistema, público ou privado”. Com isso, a Ciência Aberta possibilita novas maneiras de produzir conhecimento, objetos e sujeitos epistêmicos, e os cadernos de laboratório abertos que, por sua vez, contribuem nesse processo de inovação na tecnologia literária (artigos científicos por revisão por pares e publicados em revistas científicas). Ademais, são práticas que fortalecem e ampliam a democracia.

Recursos educacionais abertos: a Organização das Nações Unidas para a Educação, a Ciência e a Cultura (Unesco) (2015, p. 1), ao citar as diretrizes para os Recursos Educacionais Abertos (REA), define-os como “[...] materiais de suporte à educação que podem ser acessados, reutilizados, modificados e compartilhados livremente [...]”. Ou seja, são materiais ou recursos de ensino, aprendizagem e pesquisa desenvolvidos em padrões abertos, livres de barreiras de custo e acesso, e com permissão legal para uso aberto, disponibilizados à sociedade sob uma Licença ou Cessão Livre. Cabe lembrar que no ciclo de vida da REA, os recursos educativos abertos baseiam-se em resultados da investigação: descobrir, compor, adaptar, usar e partilhar; e isso fortalece e alimenta novos ciclos de produção do conhecimento.

Redes sociais científicas: têm sido aliadas no processo de divulgação da produção científica, pois, além de serem fontes e filtros de acesso à informação, elas funcionam como marketing na disseminação da presença e influência das pesquisas na *web*. São ferramentas para análise da interação entre os elementos de uma rede, assim como a função de estabelecer uma visão relacional acerca de um fenômeno, e não somente dos atributos particulares dos indivíduos, proporcionando mais visibilidade e participação social (BARBOSA NETO; ATAÍDE; ZANCAN, 2017).

Alguns autores apresentam outros pilares que dialogam com a proposta da Ciência Aberta, por exemplo: bloco de notas aberto (SILVA, 2017); reproduzibilidade – a reproduzibilidade na pesquisa é um dos princípios básicos do Método Científico, pois passa

pela transparência e disponibilização dos protocolos de pesquisa em repositórios de acesso público, contudo o “grau” de reproduzibilidade de uma pesquisa é relativo à área do estudo isso porque a similaridade almejada com replicação de experimentos são distintas nas Ciências. Todavia, a CA é primordial para que esse preceito básico do Método Científico seja atingido¹². No Brasil, por exemplo, tem-se algumas iniciativas de reproduzibilidade na pesquisa: Iniciativa Brasileira de Reproduzibilidade¹³ – financiada pelo Instituto Serrapilheira, é uma iniciativa multicêntrica para estimar a reproduzibilidade da ciência biomédica brasileira. Financiada pelo Instituto Serrapilheira; e ferramentas – para que haja a democratização científica, o uso de ferramentas tecnológicas se faz necessário para o processo de colaboração científica¹⁴.

Essas dimensões evidenciam nova práxis científica e, de certo modo, são orientadas pelos dados de pesquisas inserindo-se, portanto, nos limites da Ciência Aberta. No caso específico deste estudo, diante dos movimentos apresentados, o foco a ser desenvolvido será colocado no movimento: dados de pesquisa aberto que estão inseridos nos três grandes eixos da classificação da Ciência Aberta (Figura 2).

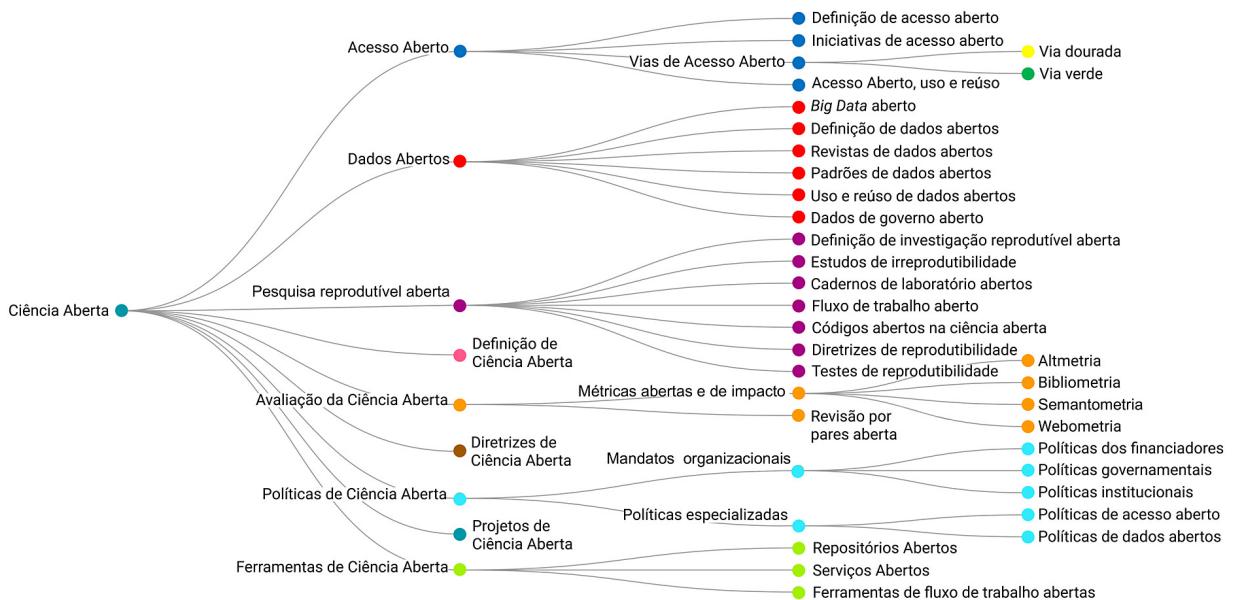
Sendo assim e diante das diversas possibilidades que o conceito de CA traz, o portal Foster – plataforma de *e-learning* que reúne os melhores recursos de treinamento, voltados para aqueles que precisam conhecer mais sobre Ciência Aberta – desenvolveu uma taxonomia que representa os principais termos relacionados à CA. Essa taxonomia busca aprofundar, ao mesmo tempo em que padroniza e estrutura a terminologia no campo da CA, e evidencia as suas múltiplas facetas. Essa classificação sistemática apresenta e reúne grandes eixos para o desenvolvimento da Ciência Aberta como acesso aberto, dados abertos e a pesquisa reproduzível.

¹² Para mais informações ver: Reproduzibilidade na ciência – Ciência Aberta USP. Acesso em: 11 jan. 2023.

¹³ Para mais informações ver: Iniciativa Brasileira de Reproduzibilidade. Acesso em: 14 nov. 2023.

¹⁴ Para mais informações ver: Ferramentas – Portal da Ciência Aberta (cienciaaberta.org). Acesso em: 6 fev. 2023.

Figura 2 - Taxonomia da Ciência Aberta¹⁵



Fonte: Ribeiro e Santos (2020)

No projeto, a CA é o termo principal que engloba nove facetas e, consequentemente, cada uma delas envolve outras ações. Desse modo, na primeira coluna, é possível identificar os movimentos que estão inseridos no contexto da ciência aberta, chamados de facetas da Ciência Aberta, quais sejam: a) acesso aberto (*open access*): estão inseridas as rotas de acesso abertos, que incluem diários e repositórios ou arquivos de acesso aberto; b) dados abertos (*open data*): incluem definição, uso e reúso dos dados, reprodução de pesquisas, diários de dados e outros; c) reproduzibilidade aberta da pesquisa: *open reproducible research* (onde estão inseridos cadernos abertos de laboratório e código aberto); d) avaliação científica aberta: *open science evaluation* (consta revisão por pares aberta, que está subordinada a métricas abertas e fator de impacto). Outras, não constam na primeira coluna, mas estão subordinadas a alguma faceta, são elas: código-aberto; redes sociais científicas; ciência cidadã; e recursos educacionais abertos. E, para completar a primeira coluna, ainda se encontram a definição de ciência aberta (*open science definition*); as diretrizes de ciência aberta (*open science guidelines*); as políticas de ciência aberta (*open science policies*); os projetos de ciência aberta (*open science projects*); e as ferramentas de ciência aberta (*open*

¹⁵ A figura apresenta um versão em português dos autores Ribeiro e Santos (2020) para a *Open Science Taxonomy*, desenvolvida originalmente pelo grupo Facilitate Open Science Training for European Research (Foster) fosteropencience.eu.

science tools) – incluindo repositório de dados.

Com base no exposto, Ciência Aberta pode se referir a inúmeras facetas, mas todas convergem para a sua essência etimológica: abertura e compartilhamento.

2.1 COMUNICAÇÃO CIENTÍFICA

Falar sobre a Ciência Aberta significa ampliar e refletir as discussões a respeito do próprio campo científico, em especial, a comunicação científica. É inerente à ciência e ao pesquisador a busca por informações previamente construídas em decorrência do processo de investigação e elucidação das descobertas. A Internet facilitou esse processo ao introduzir mudanças de natureza pública e aberta na comunicação científica.

Antes da rede de computadores, o compartilhamento dos dados de pesquisas, de forma gratuita, dependia da “boa ação” de acadêmicos e pesquisadores para favorecer a investigação e o conhecimento, visando o bem público. A *web*, por meio de ferramentas e redes abertas, permitiu o progresso dos movimentos de abertura em diversos domínios, tais como iniciativas de acesso aberto e de recursos educacionais, de dados abertos, entre outros. Essas iniciativas fortalecem um grupo de artefatos informacionais e potencializam novos ciclos de produção do conhecimento que, por sua vez, ocorre mediante evolução no processo de comunicação.

O impacto que a Internet trouxe para os pesquisadores estabeleceu novos padrões na produção, disseminação e uso dos resultados da investigação científica. Esses padrões de comunicação transcendem, muitas vezes, as barreiras financeiras, técnicas e legais que eram impostas aos pesquisadores nos séculos anteriores.

O movimento de acesso aberto trouxe reflexões a respeito das novas formas de se comunicar, visando ter maior efetividade e transparência. Isso porque o que se publica ainda está longe de serem os dados completos da produção do conhecimento científico. Abrir os dados de pesquisa vai de encontro ao *paywall* ou “muro de pagamento” – sistema de cobrança pago pelos usuários para ter acesso aos conteúdos digitais de portais das grandes editoras, das megaplatformas de artigos científicos e de algumas sociedades científicas, total ou parcial. A Ciência Aberta, portanto, envolve mudanças em todo o sistema de produção e circulação de dados e resultados de pesquisa, por isso é importante a viabilidade dessas práticas e sua compreensão.

Para Packer e Santos (2019a, p. 19, grifo dos autores),

o aspecto central da almejada qualificação de “aberta” para a ciência é o da

capacidade de comunicação permeando todos os processos e conteúdos do ciclo da pesquisa, o que, vale repetir, resgata e promove a natureza colaborativa do empreendimento científico na criação de novos conhecimentos.

Conforme os autores, a abertura dos dados consiste no conhecimento científico como bem público global e inclusivo; isto é, a formulação e a execução dos projetos de pesquisa devem contemplar a gestão do registro de um modo geral, com o intuito de disponibilizar, de forma pública e legível, aos humanos e às máquinas. Nessa perspectiva, esse registro garante a integridade da comunicação dos resultados.

Ainda de acordo com Packer e Santos (2019a, p. 20),

[...] a capacidade de comunicação em todo o ciclo da pesquisa se traduz em transparência e exposição da execução da pesquisa a controles de qualidade automatizados, de pares e do público no que diz respeito ao rigor metodológico e a padrões éticos. Tudo isso contribui para minimizar erros e más condutas, tais como plágio, conflitos de interesse, dados falsificados, comprometimento da avaliação por pares e falsas atribuições de autorias.

Essa afirmação evidencia um novo sistema de comunicação e informação mundial, inserido pela Sociedade em Rede. Tal sistema se desdobra em uma linguagem que traz reflexões de novas maneiras de se produzir e acessar conteúdo científico, colocando em destaque conceitos como participação, colaboração, compartilhamento e convergência. Antes de tudo, é preciso pensar em algumas interfaces que possibilitam essa comunicação atual:

- a) homem-máquina: atua como uma espécie de mediador entre as duas partes (homem e máquina), tornando uma sensível para a outra. Para Steven Johnson (2001), trata-se de uma relação semântica, caracterizada por significado e expressão, não por força física.
- b) espaço-informação: em *Cultura da Interface*, Johnson (2001) afirma e constata os vícios que se têm ao tratar a informação em termos espaciais nos computadores, por meio de metáforas do mundo real, denominadas metáforas espaciais. Por exemplo, quando se armazena um documento no computador, nomeia-se o arquivo (primeira metáfora espacial), salvando-o em uma pasta. Posteriormente, ao acessar esse arquivo, “[...] fingimos para nós mesmos que estamos lembrando ‘onde’ pusemos o arquivo, mas o que estamos lembrando realmente é o nome da pasta que o contém [...]” (JOHNSON, 2001, p. 61, grifo do autor). Em outras palavras, ele mostra que a interface espaço-informação apresenta uma nova forma de organização e que, na viabilidade do movimento da Ciência Aberta, trabalhar com informações em

termos espaciais sugere reorganizar tal informação, a qual é fundamental para a transparência e o compartilhamento dos dados científicos.

Na verdade, o que Johnson demonstra nada mais é do que uma ideia já apresentada por Vannevar Bush (1945) em seu Memex¹⁶ – uma máquina visionária que auxilia a memória e guarda conhecimentos. A concepção do Memex possibilita refletir sobre duas vertentes: a necessidade e a possibilidade de o leitor fazer anotações relativas ao texto durante o processo de leitura; a referência ao leitor ativo, que pode elaborar observações a determinado texto. Para Bush, o usuário constrói sua trilha de interesse, mas não consegue, sem um recurso extra – como anotar em outro espaço – realizar novamente o mesmo percurso. Ou seja, os ambientes digitais são navegáveis e possibilitam um banco de dados científicos capaz de oferecer um espaço no qual se pode transitar.

Percebe-se, com isso, um esforço dos cientistas em criar formas para divulgação do conhecimento produzido, desde os moldes tradicionais (periódicos, bibliotecas, entre outros) aos moldes atuais (por meio das redes digitais de comunicação). Contudo, ao pensar no conhecimento científico enquanto necessidade de uma população, vislumbra-se participação mais incisiva dos cientistas nos movimentos de colaboração e transparência dos dados brutos de pesquisa. Quando não se trabalha a informação, da forma semântica, apresentada por Bush (1945), cria-se um colapso ao recuperá-la nos meios digitais, fazendo com que os usuários percam tempo e fiquem limitados às mesmas prerrogativas do papel.

A partir da segunda metade do século XX – período marcado por grandes transformações sociais, tecnológicas e políticas, a legitimação das Tecnologias da Informação e Comunicação acelerou e modificou a comunicação científica. Ao comparar as formas de comunicação científica tradicionais da ciência, no século XX, antes do movimento do Acesso Aberto, são nítidas as transformações na forma de avaliar a ciência e nas alternativas de avaliar impacto científico (as altimetrias), no surgimento e na popularização dos servidores de *preprints*, entre outros. No passar dos anos, observa-se que, “[...] quando há um nível maior de compartilhamento de ideias e abertura do conhecimento, o avanço da ciência é mais rápido e as sociedades tornam-se mais avançadas, ricas e democráticas.” (KON, 2013, p. 24). Para ele,

[...] o século XX testemunhou o surgimento de grandes editoras especializadas na publicação de revistas científicas. A fim de financiar os custos de seleção e editoração de artigos científicos e a distribuição global

¹⁶ Para mais informações ver: Vannevar Bush e a Concepção do memex - Untitled (gonzatto.com).

desses periódicos, passou-se a cobrar quantias significativas por cada artigo individual ou pela assinatura da revista. Bibliotecas acadêmicas de todo o mundo passaram a assinar centenas ou milhares de periódicos. Como não havia concorrência, ou seja, apenas um editor publica uma dada revista, e os cientistas faziam questão de ter acesso a todas as revistas relevantes de sua área, o preço das assinaturas começou a subir de forma significativa. Algumas assinaturas anuais pagas por bibliotecas universitárias hoje em dia chegam a custar dezenas de milhares de dólares (KON, 2013, p. 25).

Felizmente, essa situação passou a ser questionada e revistas científicas de acesso aberto surgiram, e repositórios digitais começaram a disponibilizar textos completos livres de custos. Consequentemente, as grandes editoras perceberam que precisariam se adaptar e se inserir no movimento para continuar no negócio (HITCHCOCK, 2004; TENNANT *et al.*, 2016; WANG *et al.*, 2015). Atualmente, algumas editoras incorporam um modelo no qual todos os artigos são abertos e disponibilizados em rede e os custos, em geral, realizados com financiamento oriundo de seus projetos de pesquisa. Outras permitem que o autor pague uma quantia específica a fim de que seu texto seja disponibilizado via *web* de forma livre, mesmo em uma revista fechada. Cabe ressaltar que, no que diz respeito ao produto da pesquisa financiada com dinheiro público, isso inclui o acadêmico, vinculado a uma instituição pública, por lei ou política, deve ser abertamente acessível à sociedade, justamente porque foi ela quem contribuiu com recursos para seu desenvolvimento (SILVA; SILVEIRA, 2019).

No entanto, os desafios que enfrentam a universalização do acesso englobam todo o sistema científico, porque há um embate entre a comunidade científica e o movimento de acesso em decorrência de algumas questões, como autoria colaborativa, financiamento e outros. Sobre tal fato, cabem alguns esclarecimentos. Autores como Pierre Bourdieu (1983), Derek John de Solla Price (1969), entre outros, têm estudado e reconhecido a estrutura hierárquica da comunidade científica, e a relação entre comunidade científica e comunicação científica é evidente entre eles.

Na concepção de Bourdieu (1983), ao lançar a teoria do campo científico, trata-se de um campo social equivalente a outros, em que agem as relações de forças, de interesses e de estratégias específicas; isto é, um espaço de jogo de uma luta concorrencial. Para ele,

[...] o que está em jogo especificamente nessa luta é o monopólio da autoridade científica definida, de maneira inseparável, como capacidade técnica e poder social; ou, se quisermos, o monopólio da competência científica, compreendida enquanto capacidade de falar e de agir legitimamente (isto é, de maneira autorizada e com autoridade), que é socialmente outorgada a um agente determinado (BOURDIEU, 1983, p. 122-123).

Bourdieu (1983) deixou claro que o campo científico é um lugar de competição, com estratégias de manipulação e mobilização, e a finalidade de acumular crédito científico para o controle da ciência. Nesse contexto, a noção de “colégio invisível”, desenvolvida por Robert Boyle, cientista inglês do século 17, é readaptada e ampliada como “novos colégios invisíveis” por Derek John de Solla Price (1969), de forma bem oportuna. Esse colégio trata-se, portanto, da “nata” que se constitui no cume da comunidade científica e em volta de um *front* de pesquisa. Os membros que constituem essa elite representam um grupo de poder, porque são passíveis de controlar a gestão de fundos de pesquisa e também de laboratórios – e até as novas ideias científicas –, bem como as decisões referentes às suas estratégias.

Assim, na abordagem dos autores, há uma elite (minoria) que possui autoridade ancorada em prestígio individual, conquistada por merecimento e reconhecida pelos demais. Essa comunidade científica revela um pequeno mundo estruturado, e com estrelato científico encarregado dos discursos também científicos. Ou seja, há uma hierarquia entre indivíduos que formam as comunidades científicas; da mesma forma, há hierarquia entre os diversos tipos de veículos que podem ser usados para comunicar o conhecimento científico (MUELLER, 2006).

Para Mueller (2006, p. 30), “[...] a comunidade científica não existe em um vácuo social, mas é um dos muitos grupos sociais que compõem a sociedade contemporânea, estando, portanto, sujeita às forças presentes nessa sociedade [...]. Desse modo, percebem-se interesses nos atores que compõem essa comunidade científica no progresso ou não da comunicação científica:

- a) interesses financeiros das editoras que dominam o mercado de periódicos;
- b) interesses das instituições de pesquisa e universidades que lutam por prestígio e financiamento;
- c) interesses nacionais, políticos e econômicos que buscam o desenvolvimento e prestígio nacional;
- d) e interesse pessoal dos pesquisadores, tanto daqueles que ocupam os lugares mais altos na hierarquia – e que desejam lá permanecer –, quanto daqueles em ascensão e disputam lugares mais altos, e também os marginalizados, para quem mudanças seriam, talvez, favoráveis (MUELLER, 2006, p. 31).

Com a crise dos periódicos, algumas iniciativas foram recorrentes, entre elas a criação do primeiro arquivo eletrônico *preprints*,¹⁷ montado em Los Alamos, em 1991, pelo físico Paul Ginsparg. Essa iniciativa modificou a informação científica, permitindo que os

¹⁷ Trata-se de um documento de conteúdo original e de teor científico que foi inserido em um repositório on-line sem passar pela tradicional revisão por pares.

pesquisadores criassem arquivos eletrônicos de *preprints* e *posprints* como opção para publicação direta de seus trabalhos em texto completo. A partir daí, os primeiros periódicos eletrônicos¹⁸ apareceram e, com eles, outras iniciativas de publicação eletrônica e acesso à pesquisa: periódicos científicos eletrônicos com avaliação prévia pelos pares; servidores de *e-prints* para áreas específicas – repositórios para assuntos específicos; repositórios institucionais de universidades específicas; e, autoarquivamento em páginas pessoais dos autores (BJÖRK, 2004).

Há uma diversidade de servidores de *preprints* que existem hoje e estão sendo utilizados pela comunidade científica, por exemplo: Scielo Preprints, Humanities Commons, arXiv.org, Datacite, dentre outros. Diante disso, cabe destacar que, devido à pandemia da Covid-19, algumas dessas práticas de CA foram potencializadas, ou seja, houve um uso acelerado da publicação. Muitos periódicos, com auxílio dos pesquisadores, fizeram *fast-track*, mesmo não sendo algo novo. Neste período pandêmico, o *fast-track* foi muito utilizado pelos periódicos tradicionais, e tanto estes quanto os editores responderam com um esforço acima da média (entre a submissão e a aceitação, apenas seis dias), a fim de acelerar o processo de comunicação. Para isso, houve uma série de iniciativas que criaram *pool* compartilhado de revisores rápidos, propícios a avaliar de forma célere, sem perder a qualidade dos escritos. Um esforço complementar para disponibilizar aceleradamente os resultados de pesquisas referentes à Covid-19 (BAL, 2021).

Nesse cenário, os *preprints*, diferentes do modelo tradicional dos periódicos, foram utilizados intensamente pelos pesquisadores para o compartilhamento dos resultados de pesquisa sobre o *Sars-Cov-2*, o coronavírus. Mas é preciso cautela na fonte dos dados e o padrão de qualidade desses dados para que não haja equívoco na publicação de informações, caso que ocorreu com o periódico científico *The Lancet* em 2020. Em plena pandemia da Covid-19, a revista publicou um artigo com dados equivocados no formato *preprints* afirmindo os benefícios da hidroxicloroquina e a cloroquina no tratamento de pessoas contaminadas pelo vírus. No entanto, o estudo foi alvo de críticas pela inconsistência dos dados. Por isso, o periódico precisou fazer uma retratação e pesquisadores combateram a matéria por não ter evidências científicas; além disso os dados em questão não podem ser acessados por auditores independentes, o que evidencia preocupação com a veracidade dos

¹⁸ *The Online Journal of Current Clinical Trials* (com avaliação pelos pares) foi o primeiro periódico eletrônico a ser indexado no Index Medicus.

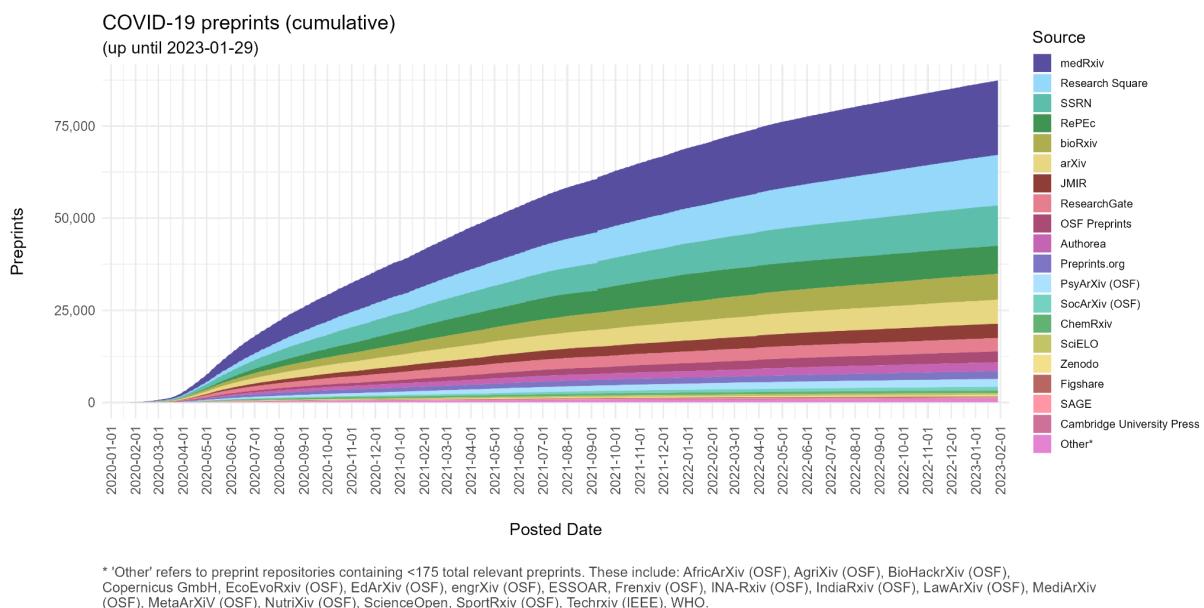
mesmos.¹⁹

Bal (2021, on-line) relata algumas preocupações sobre a duração ou permanência do impacto da comunicação e publicação na ciência:

[...] o acesso aberto, dados abertos, preprints, revisão rápida por pares e tecnologias de pesquisa de próxima geração têm desempenhado um papel fundamental na disseminação de pesquisas, mas essa experiência servirá para acelerar o movimento de pesquisa aberta? E como isso moldará a comunicação acadêmica no futuro?

Para responder a esses questionamentos de forma satisfatória dependerá do comprometimento atual das universidades, dos pesquisadores, das agências de fomentos e das bibliotecas, em relação ao processo, à gestão e às saídas de pesquisa, de maneira que possam ser mais abertas, acessíveis a todos e transparentes. A Figura 3 mostra o crescimento que ocorreu, ao longo da pandemia, da utilização dos *preprints*, a partir de janeiro de 2020, como alternativa para publicar ligeiramente as informações sobre o *Sars-coV-2*.

Figura 3 - Uso do *preprint* durante a pandemia da Covid-19



Fonte: Fraser e Kamer (2023).

Fraser e Kamer (2023) mostraram que, para todas as fontes, os *preprints* são relacionados à Covid-19, e a classificação está baseada em correspondências de palavras-chave em seus títulos ou resumos. São conjuntos de dados que estão sendo atualizados constantemente pelos pesquisadores e disponíveis no repositório do GitHub –

¹⁹ Ver: The Lancet faz retratação de estudo que apontava maior risco de morte associado à hidroxicloroquina - 04/06/2020 - Equilíbrio e Saúde - Folha (uol.com.br). Acesso em: 18 ago. 2023.

ferramenta que possibilita a colaboração e as possíveis mudanças pelos desenvolvedores em projetos compartilhados, enquanto mantêm um registro detalhado do seu progresso –, e tem por objetivo reunir informações das principais fontes de metadados de *preprints*.

Outra iniciativa que cabe lembrar, foi a contribuição de Aaron Swartz (1986-2013) para a Ciência Aberta. Swartz foi um pioneiro na internet. Além de desenvolver o Really Simple Syndication (RSS) – que organiza *feeds* da web –, o Markdown (uma linguagem simples para criar páginas da web) e o Reddit (o agora enorme site de agregação de notícias), ajudou a desenvolver a licença Creative Commons, que dá uma maneira flexível para os autores compartilharem seu trabalho, indo de encontrar a privatização da pesquisa pública e fortalecendo o princípio do acesso aberto. Esse conhecimento só foi possível graças ao Manifesto de acesso Aberto Guerrilha, criado em julho de 2008 com o intuito de “libertar a literatura científica”. Em suas palavras:

Informação é poder. Mas, como todo poder, há aqueles que querem mantê-lo para si mesmos. Todo o patrimônio científico e cultural do mundo, publicado ao longo de séculos em livros e revistas, está sendo cada vez mais digitalizado e trancado por um punhado de corporações privadas. [...] Há aqueles que lutam para mudar isso. O Movimento de Acesso Aberto tem lutado bravamente para garantir que os cientistas não assinem seus direitos autorais, mas sim garantir que seu trabalho seja publicado na Internet, sob termos que permitam que qualquer pessoa o acesse. Mas, mesmo nos melhores cenários, seu trabalho só se aplicará a coisas publicadas no futuro. Tudo até agora terá sido perdido (SWART, 2008, tradução livre).²⁰

Como defensor do acesso aberto, Aaron Swartz (1986-2013), ativista e inovador tecnológico, foi condenado e preso ao usar uma conta assinada pelo Massachusetts Institute of Technology (MIT) e baixar artigos no banco de dados JSTOR²¹. Após devolver os artigos, a JSTOR retirou a queixa, mas o processo continuou e Swartz viveu um pesadelo jurídico, principalmente, porque não conseguiu liberar os arquivos para domínio público. Nessa perspectiva, ele entendeu que houve uma falha da sua parte. Por consequência, foi acusado de crime de invasão de computadores e práticas de fraude de transmissão eletrônica, obtendo a pena máxima. Ele não aceitou e cometeu “suicídio”, em 2013. A sua morte repercutiu mundialmente e motivou outros a agir, já que, simbolicamente, ele tornou-se um mártir do

²⁰ Information is power. But like all power, there are those who want to keep it for themselves. The world's entire scientific and cultural heritage, published over centuries in books and journals, is increasingly being digitized and locked up by a handful of private corporations. [...] There are those struggling to change this. The Open Access Movement has fought valiantly to ensure that scientists do not sign their copyrights away but instead ensure their work is published on the Internet, under terms that allow anyone to access it. But even under the best scenarios, their work will only apply to things published in the future. Everything up until now will have been lost (SWART, 2008).

²¹ Abreviação de *Journal Storage*.

acesso aberto e as manifestações póstumas possibilitaram a criação do repositório Sci-Hub.

A ascensão do repositório Sci-Hub (desenvolvido por Alexandra Elbakyan) é consequência obviamente da proposta de Swart. Esse repositório abriga hoje milhões de artigos de periódicos científicos disponíveis gratuitamente para qualquer pessoa, o que facilita o acesso rápido e universal à informação científica, mas que fica prejudicada pela legislação vigente dos direitos autorais (CUNHA, 2020). Para conseguir o sucesso do repositório Sci-Hub, Elbakyan não repetiu os erros de Swart: quanto a localização (Swart foi fisicamente ao campus e conectou seu computador a um servidor do MIT, já Elbakyan morava no Cazaquistão, um país que não aplica os direitos autorais dos EUA); e quanto ao modo de liberação das informações (Swart liberava informações de dentro dos EUA, que nada mais é que o centro de informações de direitos da propriedade).

A ascensão do Sci-Hub tem gerado prejuízos financeiros para editores o que fez que os mesmos processassem nos tribunais dos EUA a criadora do repositório, no entanto todos os processos não vingaram. O sucesso do Sci-Hub se deu não apenas pela gratuidade dos artigos e fácil acesso, mas também pela descentralização do acesso, ou seja, não necessita de nenhum portal universitário único para acesso a material pago, mas é necessário usar credenciais de universidades, impossibilitando o corte ao acesso (FIX, 2019).

É preciso saber que a maioria dos artigos em acesso aberto tem uma licença Creative Commons (CC), que explica como outras pessoas podem utilizá-los. Mesmo assim, o movimento da CA enfrenta ainda muita resistência das editoras comerciais, que buscam um modelo de negócios com retorno financeiro igual ou maior do que obtêm com o acesso restrito por assinaturas (PACKER; SANTOS, 2019a, 2019b). Segundo esses autores, para resistir ao avanço da CA, a maioria das editoras se assegura do poder que exercem nas comunidades científicas, quando os seus artigos são bem ranqueados pelos indicadores bibliométricos, pois estes são comercializados pelas editoras. Contudo, em decorrência da espera pela publicação de artigos de resultados de pesquisas, o fluxo de comunicação com os manuscritos disponibilizados como *preprints*, em acesso aberto – antes ou em paralelo à submissão a periódicos – ganham força.

A construção de narrativas em ambientes digitais requer conhecimento prévio dos seus usuários e dos profissionais que alimentam esses ambientes. Dessa maneira, Murray (2003) afirma que é de competência do autor pensar em roteiros possíveis, didáticos, organizados o suficiente para serem de fácil assimilação e correspondidos, porém flexíveis de modo que

abranjam maior variedade de comportamentos humanos. Para Murray (2003), os ambientes digitais são navegáveis e, por isso, a ideia do espaço-informação é fortalecida por meio dos sistemas de realidade virtual. Assim, essa distribuição no espaço digital precisa ser atraente (no sentido semântico) e participativa; ou seja, a leitura de conteúdos científicos deve ser direcionada à ação e a CA proporciona tal atitude.

Nessa perspectiva, a comunicação científica, no contexto da Ciência Aberta, protagoniza mudanças na forma como a pesquisa é planejada, realizada, avaliada, comunicada e compartilhada, destacando não apenas o papel da ciência em prol do desenvolvimento e progresso tecnológico, mas também o papel do pesquisador e a colaboração entre os cientistas. Cabem algumas indagações: que colaboração seria essa? De que forma? Para que/quem é feita? São questões de reflexões que competem aos cientistas as respostas; contudo, um ponto seria relevante abordar – a colaboração entre eles fortalece a democracia e ao mesmo tempo põe em xeque o direito do autor no que se refere à autoria do trabalho coletivo –, e essa talvez seja a razão de muitos pesquisadores colaborarem ou não com os outros cientistas nas atividades de pesquisas.

A cultura compartilhada de resultados de pesquisas envolve responsabilidade, transparência, ética e garantia de reproduzibilidade. Essas características exigem um esforço coletivo em busca de efetividade na comunicação científica aberta e transparente, ações que fazem diferença num país corrupto e desigual. Os cientistas têm os meios, os motivos e a oportunidade de formarem uma cultura científica mais colaborativa e mais democrática.

2.2 CIÊNCIA ABERTA NO MUNDO E NO BRASIL

O movimento de acesso aberto às publicações científicas, iniciado no final da década de 1980 e começo da década de 1990, demonstra inclusão e amplia as diretrizes declaradas por instituições internacionais e nacionais para abertura dos dados de pesquisa, reconhecendo o potencial informational contido na ideia de Ciência Aberta e colaborativa. A elaboração da recomendação da Unesco a respeito da Ciência Aberta, por exemplo, baseia-se em um processo transparente e consultivo, envolvendo todos os países e partes interessadas. Trata-se de recomendações que indicam medidas concretas “[...] sobre acesso aberto e dados abertos, com propostas de ação para aproximar os cidadãos da ciência e definir compromissos para melhorar a distribuição e a produção científica em todo o mundo” (UNESCO, [2020?], on-line).

Há inúmeras vantagens para inserção nos movimentos da Ciência Aberta, dentre as quais a Unesco ([2020?], on-line) cita:

- a) maior disponibilidade e acessibilidade dos resultados de pesquisas científicas financiadas publicamente;
- b) possibilidade de rigorosos processos de revisão por pares;
- c) mais reproduzibilidade e mais transparência dos trabalhos científicos;
- d) maior impacto da pesquisa científica.

Mundialmente, vários países já aderiram a esse movimento que conclama por maior colaboração na pesquisa local e mundial e mais ampliação na abertura dos dados científicos. A saber, há algumas declarações de princípios históricos que fortalecem e promovem o movimento da CA, e tentam regularizar licenças e a abertura de dados científicos e literatura científica, por exemplo: a Iniciativa de Acesso Aberto de Budapeste, de 2001 (mas divulgada ao público em 2002); os Princípios Panton²² (Cambridge, Inglaterra, lançado oficialmente em 2010); assim como novas declarações – a Chamada para Ação sobre Ciência Aberta de Amsterdã²³ apresentada à Presidência Holandesa do Conselho da União Europeia. Esse documento defende o acesso aberto total das publicações científicas, e enfatiza um ambiente em que compartilhamento de dados e administração é a abordagem padrão para todas as pesquisas financiadas publicamente. Ademais, cita-se o atual movimento de CA na região da América Latina e Caribe²⁴, dentro do qual está a Declaração do Panamá sobre Ciência Aberta, de outubro de 2018 – documento que permitiu a redação e a abertura para sugestões e comentários. Sabe-se que essa região é composta por países com diferentes potencialidades, oportunidades e dificuldades, mas que abarcou e inovou na metodologia para a redação do referido documento.

Com isso, agências de fomento, entidades acadêmicas e sociedades científicas têm promovido ações e políticas que impulsionem práticas de CA em todo o mundo. Essas práticas são associadas a parcerias internacionais em projetos de pesquisa, e também ao acesso, dados e aos procedimentos, como é o caso da França. Outros países pioneiros na abertura de seus dados são Estados Unidos, Austrália e Reino Unido. Ressalta-se também o programa Horizon Europa, de financiamento à pesquisa 2021-2027, em substituição ao

²² A respeito ver: https://en.wikipedia.org/wiki/Panton_Principles

²³ Dados retirados do site:

<https://www.government.nl/documents/reports/2016/04/04/amsterdam-call-for-action-on-open-science>

²⁴ Disponível em:

<http://www.unesco.org/new/en/communication-and-information/portals-and-platforms/goap/access-by-region/latin-america-and-the-caribbean/>

Horizon 2020, que se baseia em três fundamentos: ciência aberta, desafios globais e inovação aberta. Esses pilares são norteadores para que países europeus possam lançar programas em Ciência Aberta a fim de atender aos futuros editais, fato que já vem acontecendo na Alemanha, com a criação de institutos; por exemplo, a Helmholtz Association. Cabe lembrar a European Open Science Cloud (EOSC) – uma iniciativa da União Europeia, onde está implantando uma rede internacional de repositórios de dados científicos abertos, que vem se tornando um dos pilares na CA europeia.

Compreendendo o contexto e a finalidade da Ciência Aberta, parte-se do princípio de que “[...] resultados de pesquisa financiada por verbas públicas são um bem público e como tal devem ser disseminados em larga escala, resguardados os aspectos éticos, de privacidade, segurança e relativos à proteção de propriedade intelectual.” (FAPESP, [202-?], on-line). Uma universidade pública – como o próprio nome diz – é de propriedade pública e recebe fundos públicos para o seu funcionamento. Assim, toda e qualquer pesquisa realizada no âmbito desse espaço deve ser de conhecimento público, com exceção daqueles protegidos por lei. Nesse contexto, no Brasil, as discussões a respeito da CA, embora sejam insuficientes, já se tornaram uma realidade.

A fim de cumprir as condições das agências de fomento e das revistas científicas, registram-se, no Brasil, várias iniciativas em Ciência Aberta de instituições públicas e privadas de ensino e pesquisa. Essas iniciativas representam um novo enfoque para o trabalho científico; para tanto, o Governo brasileiro firmou compromisso alinhado aos princípios do Governo Aberto (transparência, *accountability*, participação cidadã e tecnologia e inovação), os quais relacionam-se com os princípios da Ciência Aberta, ao lançar o Plano de Ação Nacional em Governo Aberto. Esse plano visa estabelecer mecanismos de governança de dados científicos para o avanço dela no país. Com essa iniciativa, o Brasil se uniu a outros países que buscam difundir e incentivar práticas governamentais relacionadas à transparência dos governos, ao acesso à informação pública e à participação social. Essas ações foram lançadas, em 2011, pela Open Government Partnership (OGP) ou Parceria para Governo Aberto e, como parte integrante desse movimento, o governo brasileiro endossou a Declaração e Governo Aberto, iniciando a construção do Plano de Ação Nacional.

A OGP, faz parte dos compromissos assumidos pelo Brasil desde 2011, quando foi construído o 1º Plano de Ação Nacional, e o país se tornou membro cofundador dessa parceria, assumindo internacionalmente um compromisso de estruturação dos passos iniciais

relativos a uma política de dados abertos. Trata-se de uma iniciativa multilateral internacional para fortalecer os princípios do governo aberto, capacitar o cidadão e torná-lo partícipe do governo; ademais, combater a corrupção e aproveitar as tecnologias e inovação para fortalecer a governança. Além da finalidade proposta, foi nesse período que instituiu o Comitê Interministerial Governo Aberto (CIGA) – responsável por orientar a elaboração, a implementação, o monitoramento e a avaliação do Plano de Ação Nacional sobre Governo Aberto, e identificar ações de pesquisa e desenvolvimento necessárias no âmbito desse Plano.

Os planos de ação possuem duração de até dois anos, e nesse período os governos necessitam publicar um relatório de autoavaliação a respeito da execução dos compromissos assumidos:

- a) 1º Plano de Ação Brasileiro na Parceria para Governo Aberto, 2011-2013 – firmaram-se 32 compromissos, dos quais apenas 2 foram considerados como “não cumpridos”, sendo eles: realização de pesquisa para identificação das demandas da sociedade sobre acesso à informação, visando ao aperfeiçoamento da política de transparência ativa; e elaboração e oferta de cursos de capacitação de servidores públicos em temas referentes à gestão da informação (PARCERIA PARA GOVERNO ABERTO, 2013).
- b) 2º Plano de Ação Brasileiro na Parceria para Governo Aberto, desenvolvido conjuntamente pelo Governo Federal e por diversos setores da sociedade civil, 2014-2016 – dos 52 compromissos pactuados, 36 foram integralmente implementados e 16 ainda se encontram com grau de cumprimento substantivo ou limitado (PARCERIA PARA GOVERNO ABERTO, 2016a).
- c) 3º Plano de Ação Brasileiro na Parceria para Governo Aberto, 2016-2018 – contém 16 compromissos construídos com a participação de entes subnacionais, e também dos Poderes Legislativo e Judiciário (PARCERIA PARA GOVERNO ABERTO, 2016b).
- d) 4º Plano de Ação Brasileiro na Parceria para Governo Aberto, 2018-2021 – nesse plano foram pactuados 11 compromissos. Contudo, os dados preliminares mostram que, dentre os compromissos brasileiros estabelecidos, destacam-se o compromisso nº 2, que se trata do ecossistema de dados abertos. Esse compromisso visa estabelecer, de forma colaborativa, um modelo de referência de política de dados abertos que promova integração, capacitação e sensibilização entre sociedade e as três esferas de governo, a partir do

mapeamento das demandas sociais. A elaboração desse modelo de referência para abertura de dados caracteriza-se como o mais importante, pois consiste em integrar, capacitar e sensibilizar a sociedade e as três esferas de governo no contexto da abertura de dados públicos. E o compromisso de nº 3: Inovação e Governo Aberto na Ciência indicado pela Empresa Brasileira de Pesquisa Agropecuária (Embrapa) visa “estabelecer mecanismos de governança e dados científicos para o avanço da Ciência Aberta no Brasil”. Esse compromisso, para esta pesquisa, é o mais importante, pois pretende avançar nos processos relacionados à disponibilização de dados abertos de pesquisa científica, por meio do aprimoramento de instrumentos de governança. Assim, pretende-se fortalecer a pesquisa científica, melhorar as capacidades tecnológicas de setores industriais em todos os países, incentivando a inovação e aumentando substancialmente o número de trabalhadores de pesquisa e desenvolvimento por milhão de pessoas, bem como os gastos públicos e privados em pesquisa e desenvolvimento (PARCERIA PARA GOVERNO ABERTO, 2018).

- e) 5º Plano de Ação Brasileiro na Parceria para Governo Aberto, 2021-2023 – o processo de elaboração do 5º Plano de Ação brasileiro, no âmbito da OGP, foi realizado com a coparticipação do governo e da sociedade civil entre os meses de janeiro e dezembro de 2021. Dividiu-se em três etapas: definição dos temas, oficinas de cocriação e aprovação do plano. Após os cumprimentos das etapas estabelecidas o relatório foi encaminhado a todos os órgãos de governo, para conhecimento e manifestação.

Os compromissos expostos demonstram a convergência entre Governo Aberto e Ciência Aberta, visto que a CA objetiva levar a pesquisa para a via da transparência com o intuito de se ter maior rastreabilidade, acessibilidade, verificabilidade e reproduzibilidade. Consequentemente, há evolução da produtividade científica, aumento na produção da inovação e retorno social.

Nesse sentido, uma variedade de programas institucionais vem praticando e incentivando ações associadas ao movimento global de acesso aberto às informações científicas. Atualmente, no Brasil, há inúmeras políticas e ações que realizam programas relacionados tanto com a Via Verde quanto a Via Dourada, principalmente em universidades federais e em institutos de pesquisa. Na Fundação de Amparo à Pesquisa do Estado de São

Paulo (Fapesp)²⁵, uma das iniciativas foi a Scientific Electronic Library Online (SciELO). Pioneira na adoção do acesso aberto, a SciELO se consolidou progressiva e globalmente como um dos mais importantes programas de comunicação científica em acesso aberto; ademais, é um dos exemplos de maior sucesso na área de acesso aberto por ter todas as publicações hospedadas em seu *site* distribuídas sob uma licença CC. Além dela, citam-se a Fundação Oswaldo Cruz (Fiocruz), a Empresa Brasileira de Pesquisa Agropecuária (Embrapa), o Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict), a Biblioteca Digital Brasileira de Teses e Dissertações (BDTD), a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes), o Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), a Rede Nacional de Ensino e Pesquisa (RNP), Open Knowledge Brasil, entre outros.

No que diz respeito às universidades, o *ranking* realizado pela Universidade de Leiden, em 2022, apresenta as principais universidades do mundo que mais possibilitam abertura nas informações científicas. O *ranking* Leiden, comparado a outros *rankings* universitários, oferece indicadores bibliométricos mais avançados, e representa uma perspectiva multidimensional sobre o desempenho universitário. Além disso, possibilitam o cruzamento dessas informações, tais como: impacto científico (liderada pela Universidade de Harvard com 35.050 publicações; e melhor colocada no Brasil foi a Universidade de São Paulo na 12^a com 19.552 publicações); publicação de acesso aberto (1^a Universidade de Harvard com 82.693 publicações; no Brasil, na 12^a foi a Universidade de São Paulo com 41.325 publicações); colaboração (1^a Universidade de Harvard com 86.342 publicações; no Brasil foi a Universidade de São Paulo na 12^a com 41.325 publicações); e, diversidade de gênero (Universidade de Harvard com 239.964 publicações; e melhor colocada no Brasil foi a Universidade de São Paulo na 3^a com 125.413 publicações); a UFBA obteve a 765^a posição com 3.943 publicações.

Ao comparar universidades, o desempenho pode ser visto a partir de uma perspectiva absoluta ou relativa – por exemplo, o número *versus* o percentual de publicações altamente citadas. Nesse caso, todos os indicadores precisam ser levados em consideração para nível de interpretação e tomada de decisão. Em relação a esta tese, interessam a classificação e a evidência das universidades brasileiras no que se refere ao indicador “publicação de acesso aberto”. No âmbito global, do total de 1.318 universidades, a Universidade de Harvard, Estados Unidos, a University of the Chinese Academy of Sciences e a Shanghai Jiao Tong

²⁵ Para mais informações ver: <https://www.fapesp.br/openscience/>

University (STJU), ambas na China, obtiveram as melhores classificações. Quanto às universidades brasileiras, a Universidade de São Paulo (USP), seguida da Universidade Estadual Paulista Júlio de Mesquita (UNESP) foram as melhores pontuadas. A Universidade Federal da Bahia (UFBA) ocupou a 765^a posição; as demais, a distribuição é a seguinte (Quadro 3):

Quadro 3 - Ranking das Universidades brasileiras no período de 2017-2020 no quesito publicação de acesso aberto

POSIÇÃO	UNIVERSIDADE	SIGLA	Nº DE PUBLICAÇÕES
12 ^a	Universidade de São Paulo	USP	41.325
174 ^a	Universidade Estadual Paulista Júlio de Mesquita	Unesp	15.434
188 ^a	Universidade de Estadual de Campinas	Unicamp	14.642
241 ^a	Universidade Federal do Rio de Janeiro	UFRJ	12.362
260 ^a	Universidade Federal do Rio Grande do Sul	UFRGS	11.566
285 ^a	Universidade Federal de Minas Gerais	UFMG	10.685
414 ^a	Universidade Federal de São Paulo	Unifesp	7.794
468 ^a	Universidade Federal de Santa Catarina	UFSC	6.968
520 ^a	Universidade Federal do Paraná	UFPR	6.242
567 ^a	Universidade de Brasília	Unb	5.652
588 ^a	Universidade Federal de São Carlos	UFSCar	5.464
621 ^a	Universidade Federal de Pernambuco	UFPE	5.123
671 ^a	Universidade Federal de Fluminense	UFF	4.702
682 ^a	Universidade Federal do Ceará	UFC	4603
701 ^a	Universidade Federal de Santa Maria	UFSM	4.399
707 ^a	Universidade Federal de Viçosa	UFV	4.336
711 ^a	Universidade Federal do Rio Grande do Norte	UFRN	4.311
756 ^a	Universidade Federal de Goiás	UFG	3.986
757 ^a	Universidade do Estado do Rio de Janeiro	UERJ	3.981
765^a	Universidade Federal da Bahia	UFBA	3.943
862 ^a	Universidade Federal do ABC	UFABC	3.360
870 ^a	Universidade Federal da Paraíba	UFPB	3.326
878 ^a	Universidade Federal de Pelotas	UFPEL	3.267
912 ^a	Universidade Federal do Pará	UFPA	3.131
917 ^a	Universidade Estadual de Maringá	UEM	3.107
941 ^a	Universidade Federal de Uberlândia	UFU	2.995
977 ^a	Universidade Federal do Espírito Santo	UFES	2.857
948 ^a	Universidade Federal de Mato Grosso	UFMT	2.316
1014 ^a	Universidade Tecnológica Federal do Paraná	UTFPR	2.714
1026 ^a	Universidade Federal de Juiz de Fora	UFJF	2.667

1041 ^a	Universidade Federal de Lavras	UFLA	2.592
1055 ^a	Universidade Estadual de Londrina	UEL	2.551
1205 ^a	Universidade Federal de Sergipe	UFS	1.989
1214 ^a	Pontifícia Universidade Católica do Rio de Janeiro	PUC-Rio	1.948
1231 ^a	Pontifícia Universidade Católica do Rio Grande do Sul	PUCRS	1.901
1286 ^a	Universidade Federal do Rio Grande	FURG	1.580

Fonte: Baseado no *ranking* da CTWS.²⁶

Dessa forma, este capítulo mostra o entendimento e a prática de processos na ciência em suas diversas manifestações. De fato, o conhecimento científico é um bem público e, portanto, procura fornecer alternativas desvinculadas ao modelo tradicional de publicação. Logo, o acesso aberto a esse conhecimento é uma premissa fundamental no fortalecimento da democracia. Nessa perspectiva, mediante a abertura das informações e a criação de processos transparentes, colaborativos e participativos, é possível construir uma gestão mais comprometida com os interesses da comunidade científica brasileira. Dessarte, a Ciência Aberta deve ser um processo transparente e colaborativo, de livre acesso, e a concepção de que as ideias não são propriedade privada é um benefício da cidadania.

Embora marcado por tantas vantagens, o movimento ainda sofre alguns embates para a inserção, de fato, por parte de pesquisadores e instituições. O primeiro deles é o embate da produção aberta e as restrições impostas pelo sistema de propriedade intelectual. O direito de propriedade de autoria coletiva ainda é um argumento utilizado por muitos pesquisadores para se privar de algo que é público: os dados científicos e culturais. Segundo Nielsen (2008), para se pensar numa cultura científica aberta, é necessário vencer dois grandes desafios: ótimas ferramentas on-line e realizar mudanças culturais para que essas ferramentas sejam aceitas. O sistema de reconhecimento e validação da ciência – a preocupação com o reconhecimento do trabalho científico e, consequentemente, a publicação em periódicos científicos conceituados – é outra barreira.

Contudo, o movimento Ciência Aberta vem apresentando avanços, por exemplo, na questão dos repositórios. É comum os artigos publicados ficarem disponíveis em acesso aberto, inclusive, por determinação de agências de fomento à pesquisa. O uso de novas tecnologias de informação e comunicação – não muito diferente do segundo – refere-se a uma mudança de cultura no que concerne à abertura de dados, o que inclui estar aberto às novas ferramentas de modalidades on-line para disponibilização dos dados (HOURCADE, 2015).

²⁶ Disponível em: <https://www.leidenranking.com/ranking/2022/list>

Além do exposto, com base na literatura estudada, há outros embates ou bloqueios que dificultam a inserção do movimento da Ciência Aberta no Brasil:

[...] a ausência de uma instância de governança da Ciência Aberta em nível nacional; a falta de incentivo do compartilhamento e da abertura de dados por parte do sistema de avaliação e recompensa vigente; a ausência de infraestrutura tecnológica e capacidade de análise em suporte à Ciência Aberta; o desconhecimento e a resistência da comunidade científica brasileira sobre a Ciência Aberta; a falta de incentivo do fomento à pesquisa no que tange à Ciência Aberta e às ferramentas que a sustentam; e o insucesso na aprovação da lei brasileira de acesso aberto à informação científica (BERTIN *et al.* 2019a).

Todas essas discussões levantam tensões no campo científico, ao mesmo tempo em que trazem reflexões do que realmente representa a ciência. Nesse sentido, é oportuno trazer à tona o pensamento de Everett Rogers (1983) sobre a teoria das inovações. Para o autor, a inovação (ideia, prática ou objeto que seja percebido como novo por indivíduo ou organização), os canais de comunicação, o tempo e um sistema social (combinação de influências externas e internas) são elementos definidores que influenciam na propagação de uma nova ideia. Todos esses elementos atuam conjuntamente, visto que, para difundir algo, precisa-se de comunicação via canais ao longo do tempo entre os membros de um sistema social. Logo, inovação é a razão do acesso aberto, e no sistema social em que se difunde a inovação está a comunidade científica. A teoria de Rogers, portanto, enfatiza a maneira como as inovações são propagadas a várias partes da sociedade e as opiniões a respeito dessas inovações são aspectos relevantes no processo de disseminação.

2.2.1 Propriedade intelectual e autoria colaborativa

O movimento dos dados de pesquisa abertos faz parte desse universo de práticas para uma Ciência Aberta. Talvez esse seja o movimento que apresenta questões mais delicadas, pois trata do acesso aberto aos dados primários, aqueles que servem de base para os resultados das pesquisas. Na área científica, a prioridade da descoberta envolve ganho de prestígio e poder ao pesquisador, no que Bourdieu (2004) aponta como sendo uma luta por “capital científico”. Outra questão sensível está relacionada aos direitos de propriedade intelectual, que pode ser observada na forma de direito autoral ou por algum tipo de propriedade industrial, em particular a patente, que é “[...] um título de propriedade temporária sobre uma invenção ou modelo de utilidade, outorgado pelo Estado aos inventores ou autores ou outras pessoas físicas ou jurídicas detentoras de direitos sobre a criação [...]” (INSTITUTO NACIONAL DA PROPRIEDADE INDUSTRIAL, [201-?]).

Na verdade, no contexto de um mundo capitalista e globalizado, o acesso à informação de forma gratuita ainda é um embate. Recorrendo aos textos da Declaração Universal de Direitos Humanos, de 1948, e do Acordo Internacional de Direitos Econômicos, Sociais e Culturais, observa-se que todas as pessoas têm direito à instrução. Assim, as mudanças em curso no cenário de produção e divulgação de conhecimentos, decorrentes do movimento da Ciência Aberta, são importantes indutoras à acessibilidade e à autoria compartilhada. Nesse mundo de cibercultura, boa parte da troca de obras e conteúdos, sem dúvida, é protegida pela lei de direitos autorais vigente no Brasil – Lei nº 9.610, de 1998. Porém, a estrutura básica da referida lei não é suficiente para abranger as demandas oriundas do acesso aberto, a fim de ser mais específica da autoria colaborativa em decorrência do movimento da Ciência Aberta.

Propriedade intelectual é um conjunto que reúne, sob essa denominação, as várias formas de proteção das criações humanas. Dentre elas está o direito autoral, um ramo específico da propriedade intelectual, responsável por averiguar em que medida sobre o direito autoral incide a funcionalização social de sua propriedade, incluindo as expressões literárias, artísticas e científicas, e também as questões do acesso, dados e ciência abertos, artigos científicos, coletânea e banco de dados²⁷. No Brasil, esses direitos de proteção são regulamentados principalmente pela Lei de Direitos Autorais (LDA), Lei nº 9.610/98.

Nas palavras de Lima e Santini (2008, p. 123),

[...] na sociedade da informação, a propriedade intelectual dos bens materiais se contrapõe aos interesses comuns. A produção de escassez artificial entra em conflito com a produção colaborativa e com o compartilhamento. O sistema de copyright parece não servir mais.

Ou seja, no ambiente virtual, as limitações que a lei de direitos autorais aborda são ínfimas para englobar autoria de obras colaborativas. Nesse caso, cabe lembrar da obra *O que é um autor*, de autoria de Foucault (1969). Percebe-se, nessa obra, que os textos literários passaram a ter autores quando se tornaram contraventores e passíveis de punição; ou melhor, os textos circulavam livremente sem que houvesse alguma preocupação com sua autoria. Em contrapartida, os textos científicos deveriam ser atestados por um autor. Ainda nessa obra, Foucault tentou mostrar a dissociação entre indivíduo (escritor) e função (autor); essa questão, embora polêmica, ainda é atual.

É difícil, para os cientistas com viés tradicional, pensar que sua criação ou descobrimento não pertence a si só, mas existem vários intermediários que colaboraram entre

²⁷ Curso Ciência Aberta - Série 2: Curso 1 - Unidade 1 - Aula 1: Propriedade Intelectual & Direitos Autorais (fiocruz.br).

si. Em vista disso, Michel Foucault, em uma conferência célebre apresentada em 1969, explicou a importância da figura do autor e o surgimento dessa função, destacando a força da função-autor dentro da história do pensamento contemporâneo numa perspectiva histórica e literária.

Em seu texto *O que é o autor*, Foucault (1969, p. 278) afirma que, “[...] a função-autor é, portanto, característica do modo de existência, de circulação e de funcionamento de certos discursos no interior de uma sociedade” e que, por isso, é uma obra coletiva. Ao apresentar os principais traços característicos dessa função-autor, ele estabelece a sua forma de funcionamento dentro da sociedade. A primeira característica diz respeito ao sistema jurídico institucionalizado. Ou seja, à questão da propriedade intelectual que, para ele, é responsável por determinar e articular os discursos existentes. A segunda refere-se ao tratamento da maneira como esta função-autor é exercida, isto é, ocorrem as mudanças nessa função em épocas, nos discursos e nas formas de civilizações diferentes. A terceira, porém, relaciona-se ao que ele chama de “ser de razão” (ou autor), sendo, para ele, similar ao lugar originário da escrita – operação complexa que “monta” um potente ser de razão. A quarta corresponde à existência de egos diversos e simultâneos, não remetendo apenas a um indivíduo real, mas a “várias posições-sujeito que classes diferentes de indivíduos podem vir a ocupar” (FOUCAULT, 1969, p. 284). Com isso, Foucault (1969) sugere que é o tempo de estudar o discurso não apenas pelas transformações formais, mas também pelas modalidades de circulação e apropriação do discurso.

Na LDA, no art. 11, “autor é a pessoa física criadora de obra literária, artística ou científica” (BRASIL, 1998); assim, o autor é o titular originário de todos os direitos sobre as obras que criar. Embora complexa e aparentemente distintas, a relação entre propriedade intelectual (PI) e CA pode proporcionar aumento de parcerias, diminuir custos de transação, transbordar conhecimento, divulgar de imediato o conhecimento, e isso levaria a mais parcerias e mais impulso ao desenvolvimento do conhecimento, entre outros benefícios²⁸.

Nessa perspectiva, em todas as iniciativas do movimento da Ciência Aberta, percebe-se uma evolução de que os regimes tradicionais de PI de restrição e proteção máximas minimizam a inovação e também diminuem a participação daqueles com meios e poder político limitados. Uma iniciativa que possibilitou segurança para o autor foi a Creative Commons (CC) – instrumento legal que permite aos autores instituir termos sob os quais

²⁸ Curso Ciência Aberta - Série 2: Curso 1 - Unidade 2 - Aula 6 - Ciência Aberta: Possibilidade de um novo modelo (fiocruz.br)

querem compartilhar suas obras para uso, cópia, modificação, mantendo, obviamente, o direito moral como criador. Na verdade, o intuito é promover o desenvolvimento de conteúdos ditos abertos. O CC tem por base o uso da lei de direito autoral para flexibilizá-la; além do exposto, oferece uma gama de licenças de uso mais flexíveis, feitas por combinações de cláusulas escolhidas pelo autor (LIMA; SANTINI, 2008).

O Creative Commons tem sido adaptado no Brasil, e para Lima e Santini (2008, p. 125), “[...] toda vez que o autor emite uma licença Creative Commons que regule o uso de sua obra, estabelece relação contratual sobre direitos autorais”, e, no ambiente colaborativo, a propriedade intelectual dos produtos bloqueia tanto o seguimento do processo criador como a sua apropriação comum e global por usuários. As licenças criativas equivalem à demarcação da autoria (direito moral), todavia não exibe a forma das mercadorias capitalistas; ou seja, o CC cria um meio termo legal entre os direitos reservados (LIMA; SANTINI, 2008).

Essas questões demonstram a necessidade de mudanças na LDA do Brasil, e que a natureza colaborativa, compartilhada e multidisciplinar, colabora com esse novo fazer científico; logo, é a essência de uma ciência aberta, orientada para os dados de pesquisa. Albagli (2015, p. 10) afirma que “[...] os instrumentos de propriedade intelectual, em seu atual formato, já não cabem no novo paradigma [...]. Dessa maneira, o

[...] grande desafio político-ético-legal está em conduzir uma agenda global para o entendimento de que a ciência aberta altera as práticas existentes na ciência tradicional e interfere nos mecanismos regulatórios dos DPIs que garantem a propriedade intelectual aos autores e inventores (OLIVEIRA; GUIMARÃES; KOSHIYAMA, 2019, p. 668).

No Brasil, a Faculdade de Direito da Fundação Getúlio Vargas, no Rio de Janeiro, é a representante da Creative Commons. Além de traduzir as licenças, ela as adequou à legislação brasileira. São essas licenças que liberam ou restringem o uso dos dados armazenados nos repositórios e, consequentemente, traz mais vantagens para os pesquisadores. Para Silva e Silveira (2019, p. 10),

[...] perde-se mais quando se publica de maneira tradicional porque, a depender da política editorial, os editores exigem que os direitos do trabalho sejam cedidos por um período definido ou que pode, até, vir a ser vitalício. Caso seja por um período vitalício, dependendo do contrato, há casos em que o pesquisador-autor do conteúdo precisa comprar sua obra para utilizá-la. O conteúdo em acesso aberto pode ser protegido por meio das licenças Creative Commons ou Science Commons, que permitem o ajuste dessa licença conforme o interesse do autor ou da instituição de conceder o direito de usar, distribuir, remixar as informações.

Nessa direção, o quarto paradigma não comporta mais o modelo de direito autoral vigente (Lei nº 9.610/98), pois a autoria colaborativa democratiza o acesso a dados, métodos e outros conhecimentos por intermédio das tecnologias digitais. A utilização das licenças do modelo Creative Commons é uma maneira de garantir o direito do autor e tornar o acesso às produções intelectuais por meio de licenças públicas que facilitam o acesso livre às obras por qualquer pessoa. Com isso, os seus criadores têm todo poder de decidir como seus trabalhos serão reutilizados; em outras palavras, as obras em CC estão protegidas por direitos autorais, e ao mesmo tempo os criadores compartilham legalmente seus conhecimentos e atividades profissionais à sociedade.

Dentre os tipos de Creative Commons estão: atribuição (CC BY); atribuição não comercial (CC BY-NC); atribuição sem derivações (CC BY-ND); atribuição compartilha igual (CC BY-SA); atribuição não comercial compartilha igual (CC BY-NC-SA); e atribuição não comercial sem derivações (CC BY-NC-ND). Todos os tipos de licenças apresentadas são embasadas em quatro elementos chaves: atribuição (BY) - que representa a preservação dos direitos morais do autor, requisitando que qualquer tipo de uso deverá ser acompanhado pelos devidos créditos; ND (Não a obras derivadas), ou seja, podem ser feitos downloads, cópias, redistribuição e uso, no entanto, o autor precisa autorizar qualquer modificação no material original; NC, o uso Não Comercial - não é permitido atividades com objetivo comercial; e compartilhamento sob a mesma licença ou compartilha igual (SA), ou seja, ele libera a execução, cópia, redistribuição e modificações do trabalho original. Ou seja, o fortalecimento da cultura colaborativa nos últimos anos torna as licenças CC mais relevantes e necessárias (CREATIVE COMMONS BR., 2021).

Em suma, as formas colaborativas de produção, uso e disseminação da informação requerem novos enfoques da PI e do DA, e as licenças Creative Commons são uma complementação da lei de direito autoral no Brasil, pois atendem àqueles que não querem suas restrições, já que o novo fazer científico não condiz com a legislação do direito patrimonialista (autoral). É óbvio que o sistema jurídico brasileiro não assegura todas as relações provenientes do ambiente virtual, e o CC não viola a lei de direito autoral e sim complementa, como afirma a fundação Creative Commons.

3 GESTÃO DE DADOS DE PESQUISA

As práticas da ciência contemporânea revelam um momento de preocupação com a abertura e a transparência – não apenas dos seus resultados, mas também dos dados brutos das pesquisas –, possibilitando o compartilhamento e a colaboração no trabalho científico e aumentando a eficiência e a efetividade dos recursos econômicos e sociais investidos. Nesse contexto, aumenta, entre os pesquisadores, a preocupação com os direitos de propriedade, envolvendo também os dados de pesquisa provenientes dos processos de investigações científicas desde o início da pesquisa. Essa inquietude gera uma incompreensão desse processo evolutivo – da ciência tradicional para a ciência aberta –, fortalecendo a ideia de propriedade exclusivista e individualista por parte dos pesquisadores, principalmente em universidades públicas. Consequentemente, as empresas privadas começam a obter a prioridade pelas descobertas, expandindo o sigilo e segredo comercial, ou seja, a privatização do conhecimento.

Diante da demanda cada vez mais intensa por informação, é necessário que o maior número de pesquisadores permita vias de acesso aos dados acumulados pelos seus precursores e também os atuais. Contudo, o grande volume desses dados exige auxílio no processo de organização para serem utilizados por outros usuários. Tal fato demonstra que o valor dos dados é atemporal e, muitas vezes, eles por si só não dizem muito, necessitando, portanto, de uma gestão eficaz para uso e reutilização de alta qualidade. Como afirmam Cunha e Costa (2020, p. 2), “[...] a qualidade dos dados de pesquisa e a sua proveniência tornam-se fundamentais no compartilhamento e o subsequente utilização como documento secundário.” Na concepção de Silva (2019), a disseminação dos dados de pesquisa é ponto-chave para um pesquisador, pois permite o registro, a avaliação, a disseminação e o acúmulo de conhecimento. Os dados são insumos vitais para a criação da informação e avanço da ciência, pois eles vêm também de pesquisas científicas e podem gerar um importante valor social e econômico.

Dessa maneira, neste capítulo, introduziu-se a noção dos dados nas práticas científicas. Apresentam-se conceitos acerca de dados, além de dados de pesquisa e suas diversidades quando envolvem outras áreas do conhecimento. Além disso, destaca a relevância da gestão, compartilhamento e reúso, ciclo de vidas, políticas e plano de gestão de dados de pesquisa, temas que evoluem e fortalecem o movimento da Ciência Aberta.

Visto que a gestão de dados de pesquisa tem o intuito de fornecer suporte para o

pesquisador construir o planejamento e criar um plano que possibilite a organização de seus dados, o compartilhamento com outros pesquisadores e o arquivamento para preservação e acesso a longo prazo, este capítulo descreve o processo de criação e formas de compartilhamento, sem perder de vista aspectos reflexivos sobre as percepções dos pesquisadores a respeito da abertura dos dados de pesquisa científica.

Portanto, este capítulo demonstra que esses temas ainda requerem estudos e análises mais detalhadas.

3.1 DADOS

O que são dados? Iniciar o capítulo com este questionamento parece trivial para quem é da área de Ciência da Informação. Contudo, sabe-se que não há definição única para o termo. Falar de gestão de dados de pesquisa precede conhecer termos ditos comuns no bojo da comunidade científica; é preciso conhecer o que são “dados”, “dados de pesquisa” e como geri-los.

O ato de registrar dados existe desde as civilizações antigas – como no Egito e na Mesopotâmia, entre outros –, todavia a sua evolução demonstra a necessidade de organizá-los. Historicamente, o surgimento da palavra *data* (plural da palavra *datum*, em latim) – que quer dizer “única peça de informação” – ganhou força na segunda metade do século XVII (SEMELER; PINTO, 2019). Para os referidos autores, “a palavra ‘dados’ se aplica a uma série de objetos que são originários de ambientes digitais, como textos, números, imagens, vídeos, áudios, softwares, algoritmos, equações, animações, modelos, simulações e etc”. Isto é, “dados” são os fatos, o que transcreve os fenômenos do que ocorre, no caso, em uma pesquisa científica. Nesse sentido, o uso dos computadores possibilitou um novo degrau no processo de dataficação. Trata-se de um “[...] processo que acompanha a evolução das tarefas de planejamento, previsão, reprodução e novas experimentações, até alcançar as novas formas de geração, registro e captura dos dados” (JORGE, 2018, p. 38).

Assim, no sentido matemático e estatístico, Setzer (1999, on-line) define dado como “[...] uma seqüência de símbolos quantificados ou quantificáveis [...]”. Dessa forma, um texto é um dado. Para ele, as letras são símbolos quantificados e o alfabeto constitui uma base numérica; portanto, imagens, sons e animação também são dados. Sendo assim, na perspectiva de Mayer-Schonberger e Cukie (2013), a revolução com os chamados *Big Data* – compreendidos como grandes volumes de dados ou grande banco de dados –, seu

processamento e gerenciamento possibilitam obter conclusões que não seriam possíveis com dados em pequena escala, exigindo mudanças no seu processo de preservação, gestão e análise. Dados compostos por inúmeras tipologias e combinados de forma a gerar novas análises e novos produtos. Prova disso citam-se os dados governamentais, que não são produzidos para uma pesquisa científica, mas geralmente produzidos para a avaliação de programas de governo. Consequentemente, a manipulação desses dados abertos governamentais gera pesquisas científicas, ou seja, passam a ser dados de pesquisa ou dados secundários.

A ciência sempre foi baseada em dados; contudo, o uso das tecnologias e instrumentos científicos intensificaram a sua produção. Em virtude disso, geraram-se inúmeras interpretações do conceito de dados, muitas vezes confundido como “informação” e “conhecimento”. Diante da variedade da tipologia e da dificuldade em definir dados, é comum encontrar na literatura confusões e, muitas vezes, similaridade entre esses termos, dependendo dos pontos de vista de cada autor e dos seus objetivos (SILVA, 2019).

Para iniciar, o conceito trazido pelo Decreto Federal nº 8.777 de 2016, que institui a política de dados, apresenta uma definição que atinge uma variedade de situações. Nesse decreto, entendem-se dados como uma “[...] sequência de símbolos ou valores, representados em qualquer meio, produzidos como resultado de um processo natural ou artificial” (BRASIL, 2016, on-line).

Buckland (1991, p. 102), por sua vez, traz uma perspectiva mais “concreta” de dados. Ao apresentar os paradigmas para o uso da informação, ele compara o conceito de informação como: “processo”, “conhecimento” e “coisa”. Para ele, por sua origem do latim “*datum*” – que quer dizer “única peça de informação” –, o termo “dado” é atribuído a documentos que também são considerados informação, pois se relacionam como informativos. Isto é, dado é uma informação tangível, mesmo computadorizada, pois pode ser processada e gerenciada para reutilização.

Nesse sentido informacional, o dado é um registro das propriedades de um fenômeno e é considerado o sustentáculo do quarteto hierárquico: dado (*data*), informação (*information*), conhecimento (*knowledge*) e sabedoria (*wisdom*), conhecido pela sigla DIKW, em inglês. Silva (2019, p. 22), ao afirmar que “os dados são a matéria-prima gerada a partir da concepção/coleta/observação dos fenômenos e da realidade [...]”, ratifica e fortalece a ideia da tríade, amplamente discutida na Ciência da Informação, de que, por meio da obtenção dos

dados, se concebe informação e conhecimento. Em consonância, Hjørland (2018) entende a informação como um dado processado, conhecimento, uma interpretação da informação e a sabedoria, um conhecimento interpretado.

Diante as divergências e convergências na definição de dados, optou-se, no âmbito deste estudo, por adotar a concepção de Silva (2019), sendo dado o insumo essencial para a produção da informação, ou seja, o começo das ideias. Ele ainda afirma que dado por si só não proporciona informação, é necessário interpretá-los e, por meio do seu uso e reutilização, pode-se seguir a cadeia de produção do conhecimento: dados - informação - conhecimento - informação - dados.

Segundo Silva (2019), os dados apresentam caráter quantitativo – são os qualificáveis, isto é, expressa numericamente; e qualitativo – são de caráter descritivo, ou melhor, possibilitam uma categorização. Para isso, faz-se necessário compreender a sua tipologia com base nos seguintes critérios:

- a) procedimento de coleta: dados observacionais (são dados provenientes de observações; portanto, insubstituíveis por apresentar caráter único); dados computacionais (são produtos da execução de modelos de ordenações, simulações ou fluxos de trabalho; logo, são reproduzíveis); e dados experimentais (trata-se de resultados procedentes de experimentos; portanto, fáceis de se reutilizar e de preservar).
- b) dados primários, secundários e terciários: “[...] os primários são aqueles que têm sido recolhidos por um pesquisador mediante a realização de experimentos, enquetes, entrevistas [...] que servem para dar resposta a um propósito específico [...]. Por sua vez, os secundários “estão disponíveis publicamente [...], mas têm sido recolhidos com fins não estritamente derivados do problema de pesquisa de quem se ocupam”. Já os dados terciários são uma forma de dados derivados e muito “[...] utilizados para garantir a confidencialidade dos dados primários ou secundários [...]” (SILVA, 2019, p. 29-30).
- c) segundo o grau de estruturação: dados estruturados e semiestruturados – aqueles que podem ser facilmente transferidos a outros sistemas em virtude da sua organização que segue o modelo de dados definido; e, dados não estruturados – como o próprio nome sugere, não têm um modelo de dados

definidos ou estrutura identificável corrente.

- d) dados abertos: são os que podem ser utilizados e distribuídos por qualquer pessoa, sem necessidade de autorização para o uso e reúso. A fim de identificar as permissões para esse tipo de dados, existem diversas licenças que possibilitam o uso livre (ver Cap. 2).
- e) formato de arquivo de dados: o formato e o *software* que se criam os dados, no caso de pesquisa, dependerão da maneira como os pesquisadores os coletam e analisam. Para Silva (2019), a informação digital é naturalmente dependente de um sistema operacional, colocando em perigo a obsolescência do *hardware* e *software* com o que tenham sido criados. Dessa maneira, o grande desafio da manutenção dos dados, ao longo do tempo, refere-se ao formato dos arquivos digitais. A fim de solucionar, deve-se pensar na seleção dos formatos de arquivo para o armazenamento dos dados.

Diante o exposto, comprehende-se que, para a pesquisa científica, os dados primários são essenciais no avanço da ciência; e, para os fins deste estudo, eles podem ser e serão considerados todo objeto criado em formato digital ou não, que possa ser usado como recurso primário para geração de introspecções de conhecimento. Entende-se que a comunicação dos dados é necessária tanto no avanço da ciência quanto no fortalecimento da Ciência Aberta. Cabe lembrar que a defesa do movimento da CA é que os dados utilizados nas pesquisas (brutos e/ou processados) sejam compartilhados, como “dados abertos”. No entanto, é preciso considerar as restrições cabíveis para esse compartilhamento e/ou abertura com base nas leis de proteção de dados, descritos na legislação de cada país. O foco é em dados não pessoais, ou seja, aqueles que não contêm informações sobre pessoas específicas e sem restrições de segurança nacional.

Compartilhar os dados não significa dizer que se deve abrir tudo para todos e da mesma forma. Por questões legais, pode haver restrições, e mesmo assim continuar sendo a prática de dados abertos. Dentre as maneiras existentes para o compartilhamento de dados, encontram-se: a) compartilhamento público: não há restrições para o uso; b) compartilhamento por acesso controlado: quando os pesquisadores colocam condicionantes para o acesso e a reutilização, e ainda assim continua sendo dado aberto; c) clique *sharing*: quando os pesquisadores formam consórcios e compartilham dentro desse grupo/consórcio, dificultando, dessa maneira, o engajamento e o impacto para o compartilhamento externo em decorrência das políticas restritivas; d) compartilhar por solicitação: o compartilhamento

acontece mediante requerimento. Assim, diminui o incômodo de negociar procedimentos de compartilhamento até que haja demanda; contudo, pode haver uma demora, caso ocorram várias solicitações (BYRD *et al.*, 2020).

Nesse sentido, cabe lembrar que durante o período pandêmico, a respeito de dados abertos, perceberam-se os benefícios de compartilhamento dos dados de pesquisa: de resultados epidemiológicos, das novas cepas do *Sars-CoV-2* e até de experimentos fracassados. Tal fato é importante para que outros pesquisadores ganhem tempo; embora não seja costumeiro esse tipo de compartilhamento, gera retorno imediato aos demais pesquisadores. Para isso, algumas ferramentas foram utilizadas, como Genoma Detective²⁹ – pois permite utilizar os dados brutos da máquina de sequenciamento, filtrar os resultados dos que não são vírus, montar o genoma e usá-lo para identificar o vírus. As outras ferramentas foram Share³⁰, OpenSafely³¹ – a primeira promove o compartilhamento de dados sobre financiadores, periódicos, profissionais, entre outros, e a segunda analisa os dados de registro eletrônico de saúde.

Assim, o sigilo ficou imperceptível diante a eclosão mundial por dados e informação. Além do exposto, essa crise sanitária aumentou a adoção das práticas de CA e mostrou que é plausível mudar cultura e mentalidade antigas. Houve união de pesquisadores em todo o mundo por uma causa maior: resolver ou amenizar a grave crise sanitária que atinge todos sem distinção. Isso demonstra que, independentemente das barreiras para a abertura de dados de pesquisa, é possível removê-las.

3.2 DADOS DE PESQUISA

Mundialmente, o conceito de Ciência Aberta tem sido amplamente aceito na academia devido a seus variados benefícios; além de aumentar o impacto social das suas pesquisas, economiza tempo e dinheiro e aumenta a transparência na metodologia. Embora ainda haja um hiato de conhecimento a respeito das práticas adequadas para a aplicação da CA, e sendo esta uma prática nova e complexa – quer seja, os fluxos de trabalho e/ou técnicas que podem ser usadas –, alguns pesquisadores já inseriram em seus projetos de pesquisa. Esse envolvimento e aceitação por parte de alguns pesquisadores demonstra a necessidade de mudanças de comportamento e adesão às novas práticas no que se refere à condução de pesquisas e compartilhamento de informações (PONTIKA *et al.*, 2015).

²⁹ Para mais informações ver: <https://www.genomedetective.com/app/typingtool/virus/>

³⁰ Ver: <https://covid19crc.org/research-areas/data-sharing/>

³¹ Ver: <https://www.opensafely.org/>

A Declaração de Budapeste da Iniciativa do Acesso Aberto, em 2002, foi um grande marco político, no que se refere à abertura de dados, pois estabeleceu, em nível mundial, as bases do movimento de acesso aberto. Posteriormente, em 2007, a definição de dados abertos em oito princípios fortaleceu o processo de abertura de dados e a ampliação dos usos e reúsoes da informação, impactando também na ciência. No entanto, para que um dado seja considerado aberto, significa dizer que qualquer pessoa pode acessar livremente, usar, modificar e compartilhar para qualquer finalidade desde que preserve a proveniência e sua abertura. Os princípios técnicos estabelecidos pela Open Government Working Group (Grupo de Trabalho de Governo Aberto), fortaleceram o aumento da transparência do governo, criando melhores possibilidades de controle das ações governamentais, não se limitando apenas ao combate à corrupção e ao controle de gastos, mas às qualidades das informações disponibilizadas para avaliar as políticas públicas, são eles:

- completos:** todos os dados públicos são disponibilizáveis, exceto os declarados como segredo, por motivo de segurança, ou considerados como informação privilegiada;
- primários:** os dados devem ser colhidos na sua fonte, com o maior nível de granularidade possível;
- atuais:** dados devem ser disponibilizados o mais rápido, preservando assim o seu valor;
- acessíveis:** os dados são disponíveis para o maior escopo de usuários e também para o maior número de finalidades;
- processáveis por máquinas:** a leitura dos dados deve ser facilmente obtida por máquinas que irão carregá-los e processá-los;
- acesso não discriminatório:** não há a necessidade da realização de um cadastro especial para acessá-los;
- formato não proprietário:** os dados são disponibilizados num formato não proprietário;
- livres de licenças:** nenhuma regulação de direitos autorais, patentes, propriedade intelectual ou segredo industrial deve ser necessária (WORLD WEB CONSORTIUM BRASIL, 2011, p. 17).

Além disso, o grupo afirmou que a conformidade com esses princípios precisa ser verificável e uma pessoa deve ser designada como contato responsável pelos dados. Da mesma forma, objetiva-se o aumento da transparência do governo e cria-se melhores possibilidades de controle das ações governamentais, não se limitando apenas ao combate à corrupção e ao controle de gastos, mas às qualidades das informações disponibilizadas para avaliar as políticas públicas.

Para Machado (2015, p. 208),

[...] as novas práticas de colaboração e compartilhamento da informação têm um papel relevante para a difusão de informação, cultura e conhecimento, fundamentais para o desenvolvimento humano. Na sociedade da informação,

a abertura de conhecimento humano acumulado passa a ser uma realidade possível, assumindo assim um papel prioritário em qualquer política que tenha como objetivo promover a melhoria nas condições de vida dos cidadãos e a redução das grandes desigualdades globais. Neste ponto, convergem os movimentos de defesa do acesso aberto, do acesso à informação pública, das tecnologias e protocolos abertos, da educação aberta e outros open [...].

Percebe-se, com isso, o elo existente entre acesso aberto ao conhecimento científico e dados abertos. A proposta apresentada pelo movimento da Ciência Aberta vai além do acesso às publicações; refere-se também a dados de pesquisa, às práticas de ciência cidadã, aos modelos de avaliação alternativos, entre outros que corroboram o fortalecimento da ciência e da democracia (ALBAGLI, 2015). Obviamente que a abertura e/ou compartilhamento desses dados mobilizam pontos de vista distintos entre os pesquisadores: de um lado há as iniciativas e práticas que possibilitam o acesso aos dados de pesquisas – por exemplo, os repositórios digitais de dados; em contrapartida, ainda há “[...] muita desconfiança e insegurança entre pesquisadores em torno da abertura dos dados de pesquisa” (JORGE, 2018, p. 20). Segundo a autora,

a sensibilidade e a confidencialidade de alguns dados, como aqueles que afetam a privacidade e intimidade das pessoas, mostram a necessidade de cuidado em processos de abertura. Além disso, pesquisadores ainda se sentem receosos com uma série de questões que envolvem a prioridade da descoberta, o prestígio entre os pares (capital científico), e também as questões que envolvem o patenteamento de novos produtos (JORGE, 2018, p. 20).

No âmbito internacional, países como Estados Unidos³² e alguns da Europa³³, já possuem legislações que regulamentam a abertura de dados de pesquisas financiadas com recursos públicos. Mas que tipo de dados os pesquisadores podem abrir ou compartilhar? Com base na definição de dados (Seção 3.1), a definição de dados de pesquisa também não é consenso na literatura. No Horizon 2020, a Comissão Europeia, ao estabelecer o conceito de informação científica, trouxe, em sua síntese, também o conceito de dados de pesquisa abertos: “à ‘informação científica’ refere-se a (1) artigos científicos revisados por pares (publicados em periódicos científicos) ou (2) dados de pesquisa (dados relacionados a publicações, dados com curadoria e/ou dados brutos)” (EUROPEAN COMMISSION, 2020, on-line). A Comissão Europeia ainda destacou que a informação científica gerada com financiamento público disponibilizada de forma gratuita e on-line para todos os

³² Nos Estados Unidos, a Fundação Nacional de Ciência (National Science Foundation) tem sempre um apoio ao compartilhamento de dados de pesquisas financiadas com recursos públicos. Para mais informações ver: <https://www.nsf.gov/bfa/dias/policy/dmp.jsp>. Acesso em: 9 ago. 2021.

³³ Organização para Cooperação e o Desenvolvimento Econômico (OECD) (2007, 2010, 2015), European Commission (2020) e Uk Research Councils (2011).

pesquisadores; ou melhor, ao ser financiado pelo Horizon 2020, os pesquisadores devem garantir o acesso livre a todas as publicações revisadas por pares relativas aos resultados da pesquisa.

Em consonância com a Comissão Europeia, The Engineering and Physical Sciences Research Council (2018, on-line) argumenta que dados de pesquisa são necessários para validar os resultados do trabalho científico, ou seja, “[...] material factual registrado comumente retido e aceito na comunidade científica como necessário para validar os achados da pesquisa”, tendo como benefícios de:

- a) aumentar a visibilidade da pesquisa e gerar citações, levando ao crescimento da reputação científica de pesquisadores individuais, suas equipes de pesquisa e suas instituições;
- b) reforçar a investigação científica aberta;
- c) proteger contra o uso de dados defeituosos, permitindo que os resultados publicados sejam verificados, refutados ou refinados de forma independente, melhorando assim a qualidade geral da pesquisa e incentivando a diversidade de análise e opinião e ajudando a resolver disputas científicas;
- d) estimular novas abordagens para a coleta de dados e métodos de análise;
- e) aumentar a conscientização da pesquisa em áreas relacionadas, levando a mais oportunidades de colaboração;
- f) permitir o reúso de dados para pesquisas não previstas pelos investigadores iniciais – isso aumenta a eficiência do uso do financiamento público, evitando a duplicação desnecessária da coleta de dados;
- g) permitir a criação de análises de dados mais altamente alimentadas, combinando dados de múltiplas fontes;
- h) facilitar a educação de novos pesquisadores e do público em geral.

Como visto, os dados de pesquisas não são os trabalhos científicos resultantes da pesquisa (tese, dissertação, artigo, entre outros), mas todo dado obtido e gerado para se desenvolver o produto final, que cobre uma gama de registros que podem ser estruturados e armazenados em vários tipos de arquivos (entrevistas, caderno de laboratório, questionário, dentre outros).

O documento Concordat on Open Research Data, do Reino Unido, descreve os princípios para se trabalhar com dados de pesquisa. Esse documento, que foi elaborado por

um grupo de instituições do campo científico – inclusive universidades, conselho de pesquisa e financiadores –, traz orientações e esclarecimentos de diversas instituições sobre a importância dos dados gerados na pesquisa e sua gestão, para fins primários e secundários. A respeito da gestão dos dados, Sayão e Sales (2020, p. 32), ao descreverem o conceito de dados de pesquisa, asseveram que “[...] uma boa e efetiva curadoria e gestão de dados somente se concretizará se for feita considerando as especificidades do domínio, os fluxo de pesquisas e, ainda, as características dos dados gerados nesse domínio de pesquisa”. Com isso, os autores concluem que dados de pesquisa possuem uma amplitude de significados que se transformam com base nos domínios científicos específicos, objetos de pesquisa, metodologias de geração e coleta de dados, entre outras variáveis (SAYÃO; SALES, 2020).

Ainda sobre o documento Concordat on Open Research Data, o Uk Multi-Stakeholder Group (2016) descreve dados de pesquisa como evidências que sustentam a resposta das perguntas e podem ser usados para validar os resultados, independentemente de sua forma. Além disso, mostra que esses tipos de dados podem ser informações quantitativas ou declarações qualitativas, que são aquelas coletadas por pesquisadores no decorrer de seu trabalho por experimentação, observação, modelagem, entrevista ou outros procedimentos, ou informações derivadas de evidências existentes. Também podem ser brutos ou primários ou derivados de dados primários para análise ou interpretação subsequente, ou derivados de fontes existentes, em que os direitos são mantidos por outros.

Em consonância, a Organização para Cooperação e o Desenvolvimento Econômico (OECD), em seu documento “Principles and Guidelines for Access to Research Data from Public Funding”, defende o progresso de “uma cultura de abertura e compartilhamento de dados de pesquisa” (OECD, 2007, p. 11), apontando para a necessidade de interferência e mudanças profundas na maneira como produzem, publicam, compartilham e são usados dados científicos. Um dos objetivos desse relatório é “melhorar a eficiência e eficácia do sistema científico global”. Para tanto, estabelece que os países membros considerem os princípios e diretrizes recomendados sobre o acesso a dados de pesquisa com financiamento público para instituições científicas, agências de fomento dos países membros e políticas governamentais, a fim de desenvolver políticas e boas práticas relacionados à acessibilidade, uso e gerenciamento desses dados. Assim, esse relatório apresenta dados de pesquisa como:

registros factuais (escores numéricos, registros textuais, imagens e sons) utilizados como fontes primárias para pesquisa científica e que são comumente aceitos na comunidade científica como necessário para validar resultados da pesquisa. Um conjunto de dados de pesquisa constitui uma

representação sistemática e parcial do assunto que está sendo investigado (OECD, 2007, p. 13, tradução direta).

Mesmo essa definição não abrangendo “cadernos de laboratório”, “análises preliminares e rascunhos de artigos científicos”, planos para pesquisas futuras, entre outros objetos, o acesso a todos esses produtos ou resultados de pesquisa é regido por considerações diferentes do que aqueles tratados pelo OECD. Isso porque, para essa organização, esses princípios e diretrizes são direcionados principalmente para dados de pesquisa em formato digital, legível por computador.

No âmbito nacional, a Lei nº 12.527, de 18 de novembro de 2011 – conhecida como Lei de Acesso à Informação (LAI) – foi o grande marco para abertura das informações e dados governamentais. Essa lei foi regulamentada pelo Decreto nº 8.777, de 11 de maio de 2016, que instituiu a política de dados abertos do poder executivo, além de outras normativas afins. Outra lei que merece destaque, no que se refere aos dados de pesquisa, é a Lei nº 10.973, de 2 de dezembro de 2004, que dispõe sobre incentivos à inovação e à pesquisa científica e tecnológica no ambiente produtivo e dá outras providências (BRASIL, 2004, 2011, 2016). Toda essa legislação contribui para o debate sobre a abertura e compartilhamento de dados de pesquisa entre os pesquisadores.

Em setembro de 2018, o relatório do projeto Rede de Dados de Pesquisa Brasileira (RDP Brasil), desenvolvido pela Universidade Federal do Rio Grande do Sul (UFRGS), em parceria com a Universidade Federal do Rio Grande (FURG), sob coordenação executiva da Rede Nacional de Ensino e Pesquisa (RNP) e do Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict)³⁴ – cuja finalidade foi de investigar as práticas e percepções dos usuários nacionais de serviços de Acesso Aberto a Dados de Pesquisa (AADP) –, revelou que, embora haja grande interesse pelo tema, há equívocos em relação ao que significa compartilhamento e reúso de dados de pesquisa. Além disso, comprova que a ideia de compartilhamento total e irrestrito ainda é tabu pelos respondentes, apesar da receptividade à ideia do compartilhamento e reúso de dados por parte dos pesquisadores brasileiros.

Os equívocos são nítidos, principalmente na definição do que são dados de pesquisas, usados muitas vezes como similares a dados científicos. Contudo, a literatura aponta essa discussão com um viés multidisciplinar, o que explica a complexidade na sua definição. Para Santos (2021)³⁵, há uma diferença entre os dois termos. Segundo ela, dados de pesquisas são

³⁴ Disponível em: Acesso aberto a dados de pesquisa no Brasil (ufrgs.br). Acesso em: 9 ago. 2021.

³⁵ Representante da coordenação da Rede SciELO em uma palestra proferida no dia 8 de julho de 2021 por meio do canal YouTube. Disponível em: <https://www.youtube.com/watch?v=7ewEcTATkZM>. Acesso em: 8 jul. 2021.

como dados primários na investigação científica. Eles estão relacionados a todos os registros produzidos ou usados no decurso da pesquisa, e imprescindíveis para validar os resultados da investigação; ou seja, o que é evidência da pesquisa e envolve processo de observação, experimentação ou simulação de procedimentos de pesquisa científica, algoritmo, amostras, arquivo de áudios, cadernos de laboratórios, entrevistas, questionários, entre outros. Dados científicos, porém, são bem mais amplos, referem-se a um conjunto maior de elementos que englobam dados de pesquisa. Um artigo, por exemplo, pode trazer inúmeros dados científicos, mas não dados de pesquisa.

Nessa perspectiva, diversos textos apontam e discutem dados científicos como aqueles que foram cientificamente comprovados (COSTA, 2017; SILVA, 2019), e convergem no sentido de que dados de pesquisa e dados científicos são sinônimos quando dizem respeito a dados brutos não estruturados. Porém, a terminologia dados de pesquisa remete a qualquer tipo de pesquisa – por exemplo, de mercado ou de opinião que não, necessariamente, seja científico (SALES, 2014; SAYÃO; SALES, 2015). Diante disso, para este trabalho, optou-se em utilizar o termo dados de pesquisas científicas.

Conforme Gray e colaboradores (2005), os dados de pesquisa são destaques no novo paradigma científico no que se refere à geração e ao uso de grandes volumes de dados, embora não sejam elementos no percurso da produção do conhecimento. Porém, a gestão de dados de pesquisa nunca foi tão essencial para assegurar a transparência científica, preservar dados e permitir a reutilização e a reanálise de dados, quer sejam digitais ou não digitais (TENOPIR *et al.*, 2017). Para tanto, esses dados de pesquisas “[...] precisam ser identificáveis, citáveis, visíveis, recuperáveis, interpretáveis, contextualizáveis, interoperáveis e reutilizáveis onde quesitos de consistência e procedência são considerados [...]” (SEMELER; PINTO, 2019, p. 116), e que diferem em suas tipologias e disciplinas.

Assim, é preciso analisar a natureza heterogênea e o tempo de vida dos dados de pesquisa para assegurar a qualidade do reúso desses dados, levando em consideração o grande volume (fenômeno *Big Data*) e a necessidade de práticas de armazenamento, representação e descrição dos dados de pesquisa. Cabe lembrar que o foco deste estudo é a gestão de dados de pesquisa que, por sua vez, trata-se de uma defesa amplamente discutida para a abertura desses dados, e não apenas o seu compartilhamento. Tal abertura se refere aos dados de pesquisa ainda em andamento ou não, que podem gerar novos conhecimentos, e o processo de compartilhar diz respeito à pesquisa já concluída. Logo, compreender o seu ciclo e como se

pode gerir é fundamental para conhecer os obstáculos e desafios nessa prática.

Pesquisas têm demonstrado o uso, pelos pesquisadores, de ferramentas eletrônicas para coletar, analisar e organizar dados. Porém, quando se trata de publicar seus trabalhos, não há padrões comuns e bem estabelecidos sobre como disponibilizar esses dados para outros pesquisadores. Tal fato mostra que os dados, muitas vezes, não são armazenados de forma consistente, tornando difícil encontrar conjuntos de dados associados a um artigo. Isso exige uma gestão efetiva a longo prazo, descoberta pública e recuperação. Para atender a esses requisitos, os pesquisadores precisam de ferramentas e serviços eficazes e fáceis de usar, para a curadoria, a longo prazo, de seus dados de pesquisa (AALBERSBERG; DUNHAM; KOERS, 2013; ABRAMS *et al.*, 2014).

3.2.1 Ciclo de vida dos dados de pesquisa científica

De acordo com Silva (2019), o processo científico é uma combinação de pesquisa e preservação de dados, cuja organização, com intuito de preservá-los e compartilhá-los, cumpre um ciclo de vida. Conhecer os parâmetros referentes à gestão dos dados é fundamental para o avanço científico, do conhecimento e atuação dos profissionais da Ciência da Informação. Sayão e Sales (2012) afirmam que dados de pesquisa exigem cuidados específicos em decorrência da sua fragilidade de formato e registro. Mesmo após o fim do financiamento das pesquisas, os pesquisadores precisam continuar com o tratamento dos dados, a fim de que eles possam ser analisados e reutilizados por outros, e usados futuramente como evidência.

Na concepção de Higgins (2008), o ciclo de vida dos dados pode contribuir na identificação de papéis e responsabilidades, além de possibilitar a criação de estruturas de padrões e tecnologia, e assegurar a documentação devida dos processos e políticas que envolvem o fluxo. Diante disso, Sayão e Sales (2015) elencam inúmeras concepções de modelos de ciclo de vida de dados de pesquisa: o Digital Curation Centre (DCC), o Curation Lifecycle Model, o Dataone Data Lifecycle, o Data Documentation Initiative (DDI), o Combined Lifecycle Model e o UK Data Archive Data Lifecycle. Esses modelos estruturam, de forma ilustrativa, as atividades de gestão dos dados durante sua vida, resguardando seu uso e custódia. Neles se identificam todas as fases para a gestão de dados e a responsabilidade de cada ação até a formulação e fluxos de trabalho, podendo variar conforme as políticas de cada instituição criadora.

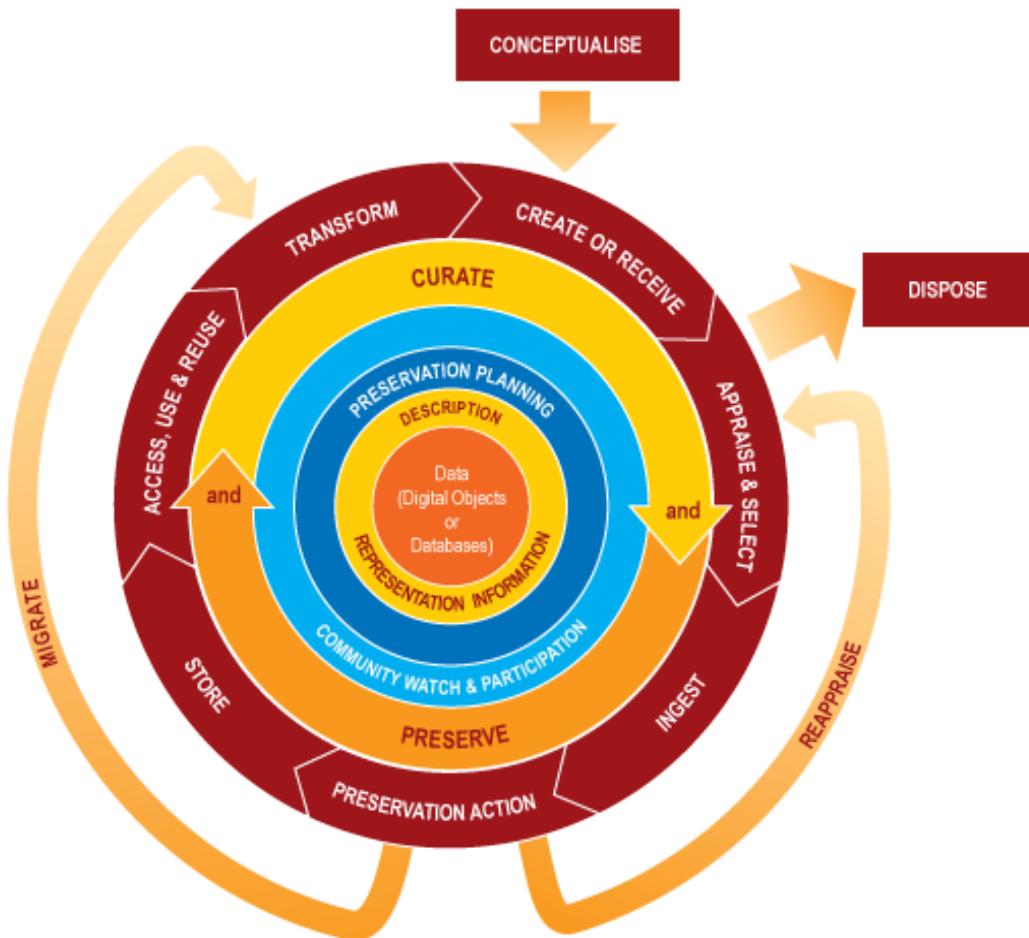
Assim, Silva (2019) revela que, geralmente, o ciclo de vida inclui fases que vão desde

a criação, passando pela aquisição, catalogação/identificação, armazenamento, preservação e chegando no acesso. Para ele, “[...] os dados em si podem ser gerados de duas maneiras: criando-os ou transformando-os a partir de conjuntos já existentes” (SILVA, 2019, p. 49). É preciso evidenciar os fatores e as diferentes fases do processo de Ciclo de Vida dos Dados de Pesquisa (CVDP), pois a vida útil desses dados ultrapassa a etapa da investigação.

Considerando que novos projetos de pesquisa podem ser gerados a partir desses dados, torna-se imprescindível a elaboração eficiente do CVPD. Por isso, vários modelos de ciclo de vida aplicados foram criados, no intuito de refletir as especificidades que demandam esse processo. Silva (2019, p. 49) afirma que “[...] a importância dos modelos de ciclo de vida é que proporcionam uma estrutura para a consideração das muitas operações que devem ser realizadas em um registro de dados durante seu uso”. Diante disso, muitos pesquisadores e profissionais da informação têm tratado o enfoque do ciclo de vida para a gestão de ativos digitais com base no Curation Lifecycle Model, desenvolvido pela Digital Curation Centre. Esse modelo trata especificamente das necessidades relacionadas com a curadoria digital, que, de certo modo, corresponde às técnicas relacionadas à *e-Science* (ALBUQUERQUE, 2018; SILVA, 2019). Segundo o *site* do DCC, a curadoria digital “envolve a manutenção, a preservação e a agregação de valor a dados de pesquisa durante o seu ciclo de vida”, e a gestão ativa desses dados reduz as ameaças ao seu valor de longo prazo e a duplicação de esforços na criação de dados de pesquisa, minimizando os riscos da obsolescência digital.

Além do exposto, percebe-se que esse modelo da Figura 4 se adequa a determinados contextos que exigem os dados de pesquisa, pois permite aos curadores identificar as fragilidades nas políticas potenciais, possibilitando que os criadores detectem falhas no processo de arquivo ou preocupações que possam ser incluídas ou eliminadas nas práticas. Também define funções e responsabilidades específicas dentro de seu projeto/instituição nas diferentes etapas.

Figura 4 - Modelo de ciclo de vida dos dados da Digital Curation Centre (DCC)



Fonte: Digital Curation Centre ([2018?]).

No centro, representado como “anéis concêntricos”, encontram-se os dados digitais que estão identificados com objetos digitais (arquivos de texto, imagem, som e outros) ou banco de dados simples e complexos, por exemplo, coleções estruturadas de registros ou dados armazenados em um sistema de computador. Em torno dos dados, o DCC vislumbra o que denomina “ações completas do ciclo de vida”, que engloba a descrição e representação da informação (*representation information*). Referem-se aos metadados administrativos, descritivos, técnicos, estruturais e de preservação, usando padrões apropriados, assegurando: a descrição e controle adequado a longo prazo; o planejamento da preservação (*preservation planning*), inclui planos para gerenciamento e administração de todas as ações de ciclo de vida; a vigilância e participação da comunidade (*community watch e participation*), mantendo vigilância sobre as atividades comunitárias apropriadas e participação no desenvolvimento e

compartilhamento de padrões, ferramentas e *software*; e ainda as ações organizar (*curate*) e preservar (*preserve*), mantendo Ciência e realizando ações administrativas planejadas para promover a curadoria e a preservação (DIGITAL CURATION CENTRE, [201?]; OUCHI, 2019).

As demais ações – como as ações sequenciais – estão divididas da seguinte forma: conceituação, criação ou coleta dos dados, avaliação e seleção, inserção, preservação, armazenamento e transformação. Conforme Silva (2019), essas ações referentes ao processo de armazenamento se apresentam da seguinte forma:

- a) *conceptualise* (conceituação): concepção e planejamento da criação de objetos digitais, inclusive a captura dos dados e opções de armazenamento;
- b) *create or receive*: criação ou coleta dos dados, em que a criação refere-se aos dados originais gerados e registrados pelos pesquisadores e a coleta aos dados obtidos de outras fontes preexistentes. Na medida em que diferentes pesquisadores gerenciem os dados, é essencial descrever as distintas normas que se devem comprovar em relação às políticas locais, atribuição de metadados arquivísticos administrativos, descritivos, estruturais e técnicos;
- c) *appraise and select* (avaliar e selecionar): nessa etapa, os pesquisadores ou curadores avaliam e selecionam os dados para a preservação a longo prazo, com base no plano de gestão, e isso envolve as políticas e os requisitos legais. Nessa fase, pode haver o descarte de dados que não sejam mais úteis, reavaliações e migração de dados para outros formatos;
- d) *ingest* (transferência): essa fase se refere à ação da preservação que condiz as seguintes atividades: controle de qualidade, catalogação, classificação, registro de metadados semânticos e estruturais, entre outras. Caso os dados estejam sem tratamento, estes devem retornar ao pesquisador para sua avaliação futura;
- e) *preservation action* (ação de preservação): está relacionada com a fase do armazenamento dos dados. Um compromisso que deve ser mantido para assegurar que os dados permaneçam seguros. Além disso, podem estar associadas com a manutenção do *hardware* de armazenamento, realização de cópias de segurança, e outros;
- f) *store* (armazenamento da informação): é a fase correspondente ao armazenamento da informação seguindo os padrões estabelecidos a tais efeitos;
- g) *access, use and reuse*: após o armazenamento dos dados de forma segura, o

acesso, uso e reutilização entram em cena. Para Silva (2019, p. 53), “[...] as ações de curadoria digital associadas a esta etapa se concentram em manter, por exemplo, o controle e a adequação dos metadados mediante interfaces personalizadas de buscas”;

- h) por fim, chega-se à etapa final do ciclo de vida, que é a *transformation* (transformação): refere-se à necessidade de transformar os dados por meio do tempo em diferentes formatos para evitar a obsolescência do *software*, ou seja, criação de novos objetos digitais a partir dos objetos originais (DIGITAL CURATION CENTRE, [201?]; SILVA, 2019).

Para finalizar o modelo, o DCC estabelece os estágios que são aplicados eventualmente, chamados de ações ocasionais. Com essas ações, é possível interromper ou reordenar as ações sequenciais como desdobramento de uma decisão; são elas:

- a) eliminação (*dispose*): eliminar os dados que não foram selecionados para curadoria e preservação de longo prazo, de acordo com políticas documentadas, diretrizes ou exigências legais;
- b) reavaliação (*reappraise*): retornar ao dado cujos procedimentos de avaliação foram falhos, para nova avaliação e possível seleção para curadoria;
- c) migração (*migrate*): migrar os dados para um formato diferente; isto pode ser feito no sentido de compatibilizá-lo com o ambiente de armazenamento ou para assegurar a imunidade do dado contra a obsolescência de *hardware* e de *software* (SAYÃO; SALES, 2012, p. 186).

O modelo da DCC define um método de idealização de gerenciamento documental baseado no ciclo de vida dos dados, com a finalidade de alinhar as tarefas de curadoria digital às etapas do ciclo. Para isso, algumas questões devem ser consideradas no processo de elaboração do CVDP: privacidade, qualidade, direitos autorais, integração, disseminação e preservação. Sayão e Sales (2012, p. 126) compartilham da ideia quando afirmam que:

[...] os elementos chaves do modelo são: dados, objetos digitais e bases de dados. No centro do ciclo de vida da curadoria está especificamente o dado digital (codificado em formato binário). A ideia objetos digitais simples, que são aqueles compostos por um único arquivo, identificador e metadados; e os objetos digitais complexos, que são formados pela combinação de outros objetos digitais, por exemplo, uma página web.

O ciclo de vida de curadoria dos dados deve ser realizado permanentemente, pois assim é possível ter uma visão ampla do alto nível dos estágios envolvidos no gerenciamento e preservação de dados para uso e reutilização. É importante destacar que, ao aplicar um modelo de ciclo de vida de dados de pesquisa, é preciso considerar o contexto cultural presente em cada instituição, ou seja, o modelo deve ser adaptado a diversas realidades

(SAYÃO; SALES, 2015). Os autores orientam aos indivíduos envolvidos que, ao se iniciar à gestão do ciclo de vida dos dados de pesquisa, é necessário conhecer as possibilidades de continuidade de trabalho com os anteriores, visto que os dados e as coleções de dados das pesquisas geralmente possuem tempo de vida maior que os projetos que as criaram, conforme mencionado anteriormente. Portanto, é necessário descrever e documentar os processos utilizados para criação dos dados.

3.2.2 Gestão e plano de dados de pesquisa

As metamorfoses provenientes da evolução da *e-Science* ou do “quarto paradigma” – que é a Ciência Aberta – colocam em evidência a necessidade de adaptação às novas práticas da ciência contemporânea, e revelam um momento de inquietação com a abertura e a transparência dos resultados e dos dados brutos das pesquisas. Neste período de transição, os atores emergentes desse processo, que são os pesquisadores, instituições acadêmicas e agências de fomento, também despertam a necessidade de armazenar, preservar, gerenciar e compartilhar os dados de pesquisa. Nesse sentido, o uso das tecnologias computacionais facilita o acesso a esses dados, muito embora seja um desafio, para muitos, armazenar e preservar nessas mídias, isto é, se os dados não estiverem organizados e comprehensíveis, tornam-se irrelevantes.

Reconhecer a importância de uma gestão adequada para os dados de pesquisa é um passo fundamental. Vários conceitos estão relacionados à gestão desses dados, e engloba todo o chamado “ciclo de vida dos dados”. Sayão e Sales (2015, p. 1) compartilham a ideia de que

[...] a gestão de dados cobre todos os aspectos relativos à manipulação, organização, documentação e agregação de valor, e tem um papel crucial como facilitador nos processos de compartilhamento dos dados, na garantia da sustentabilidade e acessibilidade dos dados em longo prazo.

Trata-se, portanto, de um processo de “cuidar” dos dados de pesquisa a fim de preservá-los ao longo do seu ciclo de vida. Essa é uma etapa fundamental para a publicação dos dados abertos, pois permite novas análises e reúso a partir dos achados do estudo original.

No mesmo sentido, Cox e Pinfield (2014, p. 300, tradução direta) argumentam que a gestão de dados de pesquisa “[...] consiste em várias atividades e processos associados ao ciclo de vida dos dados: design e criação de dados, armazenamento, segurança, preservação, recuperação, compartilhamento e reutilização”. Além disso, incluem todas as capacidades técnicas, considerações éticas, questões legais e estruturas de governança. Todas essas atividades que envolvem o fluxo e o ciclo da gestão de dados de pesquisa impulsionam

fatores como a Ciência Aberta, a demanda dos financiadores de pesquisa e editores por reconhecerem a necessidade de melhoria no acesso aos resultados das investigações (ALMEIDA, 2019).

Uma gestão adequada dos dados é essencial, mas os desafios são muitos, principalmente no que se refere à responsabilidade sobre a gestão de fato desses dados. Para Gold (2007), é preciso que haja cooperação entre os pesquisadores, as bibliotecas, as instituições e as agências financeiras, pois todos esses atores envolvidos na infraestrutura de pesquisa e comunicação científica buscam seus papéis e responsabilidades na gestão de dados. Seja quem for o responsável, sabe-se que é preciso criar políticas que abarquem a gestão de dados de pesquisa. Na esfera das políticas,

o primeiro passo poderia ser atrelar a gestão e o compartilhamento de dados como obrigação para as pesquisas com financiamento governamental. Outra necessidade incluiria a criação de padrões e práticas para citação de dados. O terceiro grande passo seria o desenvolvimento de práticas e mecanismos de controle de acesso adequado aos dados, pois em casos onde exista limitação de acesso ou restrições quanto à propriedade intelectual (ALMEIDA, 2019, p. 43-44).

Autores como Corrêa (2017) e Sayão e Sales (2015) afirmam que os principais responsáveis pelos dados de pesquisas são os pesquisadores, cabe a eles o planejamento e a sistematização dos dados no processo do ciclo de vida, desde o planejamento ao armazenamento. De fato, eles criam e detêm a coleta dos dados, mas não os fazem sozinhos, logo, a responsabilidade na garantia da qualidade, integridade e preservação compete a todas as pessoas envolvidas nesse processo. Mesmo assim, a literatura mostra uma ruptura entre o acesso e a abertura dos dados de pesquisa, o que escancara a dificuldade, por parte dos envolvidos nesse processo, em compreender as etapas necessárias para criar políticas de acesso aberto.

A fim de solucionar a lacuna existente a respeito da implementação do acesso à abertura dos dados de pesquisa – bem como os componentes que precisam ser abordados para apoiar as políticas de acesso aberto dos financiadores –, Foster especificou duplamente o conceito de Ciência Aberta por meio da taxonomia. Primeiro, mapeou o campo e apresentou os componentes principais da CA para aqueles que não estavam familiarizados com o conceito (Figura 2); depois, aproveitando a organização, forneceu uma terminologia compactada para reduzir os conflitos descritivos existentes no campo da CA e promover seu desenvolvimento mediante a consistência das práticas abertas que se inter-relacionam. Dessa maneira, a utilização da taxonomia classificou não apenas o campo temático, mas também

aproveitou a sua capacidade de organização para fornecer outra estrutura ligada à Ciência Aberta, mas que traria novas reflexões a respeito de tópicos não abordados na primeira estrutura, que são gestão de dados de pesquisa e questões legais, conforme apresentado na Figura 5:

Figura 5 - Taxonomia de gestão de dados de pesquisa



Fonte: Pontika e colaboradores (2015).

Para o Foster, gestão de dados de pesquisa compreende as disciplinas relacionadas ao gerenciamento desses dados como um recurso valioso. Tal gestão não se limita apenas a governança de dados de pesquisa, segurança e qualidade, ela abrange outros tópicos que agem conjuntamente como apoio no gerenciamento. Os referidos tópicos possuem os seguintes objetivos:

- a) um plano de gerenciamento de dados é um documento formal que descreve como lidar com dados de pesquisa durante e após a conclusão do projeto de pesquisa;
- b) políticas de gerenciamento de dados de pesquisa refere-se a um conjunto de princípios, tipicamente exigidos por vários órgãos, que devem ser seguidos ao gerenciar dados de pesquisa;
- c) serviços de gerenciamento de dados de pesquisa consulta aos serviços (on-line) que auxiliam no processo de gerenciamento;
- d) padrões de gerenciamento de dados de pesquisa;
- e) ferramentas de gerenciamento de dados de pesquisa;

A estrutura apresentada por Foster corresponde apenas às ferramentas que auxiliam no

processo de gerenciamento de dados de pesquisa, subdividida em cinco elementos (primeira coluna). No entanto, diante das possibilidades e o avanço das discussões a respeito de gestão de dados de pesquisa, percebe-se a preocupação com as questões legais por trás do compartilhamento, reúso e reprodução dos textos que envolvem a abertura desses dados. Por isso, Pontika e colaboradores (2015) apresentaram uma nova estrutura, na qual as questões legais foram consideradas em decorrência da complexidade, por meio de classificação facetada, permitindo classificar objetos usando mais de uma taxonomia, e por entender que a condução da pesquisa está relacionada a um sistema de princípios morais. Dessa maneira, as taxonomias de gestão de dados e questões legais e de pesquisa têm apenas um nível de termos mais estreitos. Os autores consideraram também o aspecto ético; no entanto, eles não desenvolveram a estrutura com esse aspecto, mas sinalizaram que será uma tarefa posterior.

O acesso aos dados de pesquisa tornou-se uma exigência mundial; consequentemente, os financiadores de pesquisa têm se preocupado em melhorar essa gestão de dados e enfatizado a relevância do compartilhamento das produções científicas (UK RESEARCH COUNCILS, 2011). Como condicionante a um aplicativo de financiamento, as agências de fomento têm exigido um plano de gestão de dados de pesquisa, delegando para as instituições a responsabilidade de gerir, compartilhar e preservar esses dados, isso quando não houver repositórios de dados de assuntos nacionais.

No Brasil, as agências de fomento são importantes instituições governamentais, fomentadoras no desenvolvimento científico do país. Os primeiros debates para se criar uma entidade governamental com a finalidade de fomentar o desenvolvimento das produções científicas ocorreram na década de 1930, havendo participação da Academia Brasileira de Letras, para a criação de um Conselho de Pesquisas. Contudo, em 1951, o Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq)³⁶ foi sancionado pelo então presidente Eurico Gaspar Dutra, cujo objetivo era promover e estimular o desenvolvimento da investigação científica e tecnológica, mediante a concessão de recursos para pesquisa, formação de pesquisadores e técnicos, cooperação com as universidades brasileiras e intercâmbio com instituições estrangeiras. Atualmente, esse Conselho desempenha papel crucial para formular e conduzir as políticas públicas de ciência, tecnologia e inovação. Suas ações são fundamentais para o desenvolvimento nacional e o reconhecimento das instituições de pesquisa e dos pesquisadores brasileiros pela comunidade científica internacional (CNPq, [20-?]). Atualmente, o CNPq é vinculado ao Ministério da Ciência, Tecnologia e Inovação

³⁶ Ver: http://portal.cnpq.br/web/guest/apresentacao_institucional/

(MCTI) e possui uma Coordenação Nacional de Aperfeiçoamento de Pessoal de Nível Superior (Capes), a fim de garantir recursos específicos de formação de cientistas e pesquisadores no ambiente acadêmico. Essa coordenação foi criada no segundo Governo de Vargas (1950-1954), devido à necessidade de formação de especialistas e pesquisadores nos diferentes ramos de atividade (MONTEIRO; LUCAS, 2019).

O CNPq é uma agência de fomento no âmbito federal. No âmbito estadual, atuam as Fundações Estaduais de Amparo à Pesquisa (FAPs) cujo intuito é promover a pesquisa em Ciência, Tecnologia e Inovação (CT&I). Há 27 FAPs que integram o Sistema Nacional de Ciência e Tecnologia, ao lado de instituições tradicionais, como o CNPq e a Financiadora de Estudos e Projetos (Finep). São entidades que exercem papel significativo nas definições da política científico-tecnológica nacional.

A primeira FAP foi fundada em 1962, em São Paulo; em 1964, foi criada a Fapergs, no estado do Rio Grande Sul; em 1980, a Faperj, no estado do Rio de Janeiro; em 1985, no estado de Minas Gerais foi fundada a Fapemig e, a partir de 1990, outras fundações foram criadas. Atualmente, as FAPs são as seguintes:

Quadro 4 - Fundações de Amparo à Pesquisa no Brasil

Fundação	Sigla	Ano de Criação
Fundação de Amparo à Pesquisa de Alagoas	Fapeal	1990
Fundação de Amparo à Pesquisa do Acre	Fapac	2012
Fundação de Amparo à Pesquisa do Estado do Amapá	Fapeap	2009
Fundação de Amparo à Pesquisa do Estado do Amazonas	Fapeam	2002
Fundação de Amparo à Pesquisa do Estado da Bahia	Fapesb	2001
Fundação Cearense de Apoio ao Desenvolvimento Científico e Tecnológico	Funcap	1990
Fundação de Apoio à Pesquisa do Distrito Federal	FAPDF	1992
Fundação de Apoio à Pesquisa e Inovação do Espírito Santo	Fapes	2004
Fundação de Apoio à Pesquisa do Estado de Goiás	Fapeg	2005
Fundação de Amparo à Pesquisa e ao Desenvolvimento Científico e Tecnológico do Maranhão	Fapema	1990
Fundação de Amparo à Pesquisa do Estado do Mato Grosso	Fapemat	1994

Fundação de Apoio ao Desenvolvimento do Ensino, Ciência e Tecnologia do Estado do Mato Grosso do Sul	Fundect	1998
Fundação de Amparo à Pesquisa do Estado de Minas Gerais	Fapemig	1985
Fundação Amazônia de Amparo a Estudos e Pesquisas do Pará	Fapespa	2007
Fundação de Apoio à Pesquisa do Estado da Paraíba	Fapesq	1992
Fundação Araucária de Apoio ao Desenvolvimento Científico e Tecnológico à Pesquisa do Estado do Paraná	FA	2000
Fundação de Amparo à Ciência e Tecnologia de Pernambuco	Facepe	1989
Fundação de Amparo à Pesquisa do Estado do Piauí	Fapepi	1993
Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro	Faperj	1980
Fundação de Amparo à Pesquisa do Estado do Rio Grande do Norte	Fapern	2003
Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul	Fapergs	1964
Fundação de Amparo ao Desenvolvimento das Ações Científicas e Tecnológicas e à Pesquisa do Estado de Rondônia	Fapero	2011
Fundação de Amparo à Pesquisa do Estado de Roraima	Faperr	2022
Fundação de Amparo à Pesquisa e Inovação do Estado de Santa Catarina	Fapesc	1990
Fundação de Amparo à Pesquisa do Estado de São Paulo	Fapesp	1962
Fundação de Apoio à Pesquisa e Inovação Tecnológica do Estado de Sergipe	Fapitec	2005
Fundação de Amparo à Pesquisa do Estado de Tocantins	Fapt	2011

Fonte: Confap ([2023]).

É necessário considerar que o desenvolvimento científico de um país é fundamental para a construção de uma sociedade alicerçada no conhecimento, auxiliando no crescimento econômico e na diminuição das desigualdades e exclusões sociais. No âmbito da União Europeia (UE), várias iniciativas estão sendo realizadas visando a implantação da CA. A Horizon 2020³⁷, por exemplo, lançada em 2014, tem por objetivo facilitar descobertas e avanços científicos, transferindo ideias inovadoras dos laboratórios para o mercado. Com quase 80 milhões de euros de orçamento, esse programa foi sucedido pelo Horizon Europa.

³⁷ Ver:

<https://wayback.archive-it.org/12090/20220124080448/https://ec.europa.eu/programmes/horizon2020/en/what-horizon-2020>

Além deste, outro programa exitoso é a Australian National Data Service (ANDS)³⁸ – financiado pelo governo australiano, a ANDS faz parceria com instituições de pesquisa e agências produtoras de dados a fim de permitir a transformação dos dados, por exemplo: não gerenciados para gerenciados, desconectados para conectados, invisíveis para encontráveis e de uso único para reutilizáveis – que permitirão que os dados de pesquisa da Austrália se tornem um recurso estratégico nacional para apoiar pesquisas melhores, mais eficientes e defensáveis e melhores contribuições políticas. Além disso, fornece conexões ricas entre dados, projetos, pesquisadores e instituições e promove a visibilidade das coleções de dados de pesquisa australianas nos mecanismos de pesquisa.

Outras iniciativas globais são: Política Nacional de Ciência Aberta (Portugal), Le Plan national pour la science ouverte (França), entre outros. Com isso, as principais instituições de financiamento da pesquisa global – Wellcome Trust, Fundação Bill e Melinda Gates e National Institutes of Health (NIH), e as agências de financiamento dos EUA como a National Science Foundation – passaram a exigir que os pesquisadores fornecem o plano de gestão de dados no ato de submissão de projetos de pesquisa. Ademais, solicitam a publicação, em acesso aberto, do conjunto de dados subjacentes aos artigos resultantes da pesquisa financiada.

A respeito disso, a Council of Library and Information Resources publicou um relatório apresentando como as instituições de pesquisa nos Estados Unidos estão respondendo aos requisitos de gestão de dados da National Science Foundation e os National Institutes of Health e outros órgãos federais. Além disso, o documento mostrou que tanto as bibliotecas quanto os profissionais da informação respondem melhor às necessidades de gestão de dados de pesquisa nas universidades. Esse relatório sugere que as instituições tomem iniciativas para enfrentar o problema ou a falta da gestão de dados, os quais podem ser inevitáveis a longo prazo, pois cabe a elas essa responsabilidade (HALBERT, 2013). A autora ainda questiona: será que as instituições estão respondendo a esse desafio?

Outra pesquisa a respeito do PGD foi desenvolvida por Akers e colaboradores (2014, tradução direta) no qual este estudo, os pesquisadores abordam a participação das bibliotecas de pesquisa acadêmica no desenvolvimento de suporte para gestão de dados de pesquisa, incluindo novos serviços e infraestrutura. Os autores apresentam histórias relativas a oito universidades diferentes que desenvolveram programas de suporte PGD, focando no papel

³⁸ Mais informações ver: <https://data.gov.au/data/organization/about/australian-national-data-service>. Disponível em: 23 out. 2022.

proeminente da biblioteca na educação e assistência aos pesquisadores, com gerenciamento de seus dados ao longo do ciclo de vida da pesquisa. Consequentemente, foi possível construir cronogramas para cada universidade que retrata passos-chave no suporte de construção de PGD, e discutir semelhanças e diferenças entre as universidades em motivação, visando fornecer suporte, colaborações entre as unidades do campus, avaliação das necessidades e serviços, e mudanças no pessoal. Eles mostram que é possível, apesar dos desafios.

No Brasil, a respeito do plano de gestão de dados de pesquisa, o tema ainda é recente, sendo poucas as iniciativas para a promoção da publicação de tais dados. Em 2017, a Fundação de Amparo à Pesquisa do Estado de São Paulo (Fapesp) passou a exigir a apresentação de um plano de gestão de dados para os projetos submetidos aos seus programas de fomento. A Fundação reconhece a importância da gestão adequada dos dados de pesquisa, considerando necessário que aqueles resultantes de projetos financiados por ela sejam gerenciados e compartilhados. Embora seja a única agência de fomento brasileira que exige o PGD, a Fapesp necessita usar a ferramenta do exterior para a criação do Plano, pois ainda não possui formulário próprio.

A Fapesp, embora reconheça a importância da gestão adequada dos dados de pesquisa como parte essencial das boas práticas de pesquisa, não tem um modelo próprio para tais plano, é a , pois, para ela, a criação do plano varia conforme o domínio do conhecimento, os tipos de dados considerados e como os responsáveis pelo projeto pretendem disponibilizá-los. Contudo, a Fundação entende que é necessária uma gestão apropriada dos dados, pois facilita a reproduzibilidade da pesquisa e permite promover novas pesquisas, graças à possibilidade de reúso e compartilhamento; além disso, ajuda a realização de novas análises, com execução de outros testes ou métodos de análise e facilita o treinamento de novas gerações de pesquisadores e a exploração de tópicos não previstos no projeto original. Para tanto, a apresentação de um PGD é um elemento obrigatório na fase de submissão de um projeto e apresenta como ferramentas on-line que criam planos de gerenciamento de dados e auxilia na promoção da pesquisa DMPTool³⁹ e a DMPonline⁴⁰, respectivamente associados a Planos de Gestão de Dados nos EUA e países do continente europeu.⁴¹

O Plano de Gestão de Dados para a Fapesp deve responder a duas perguntas básicas: 1. Quais dados serão gerados pelo projeto? e 2. Como serão preservados e disponibilizados,

³⁹ <https://dmptool.org>

⁴⁰ <https://dmponline.dcc.ac.uk>

⁴¹ Para mais informações ver: Plano de Dados Abertos.pdf (ufrrgs.br)

considerando questões éticas, legais, de confidencialidade e outras? Com isso, mundialmente, várias universidades e organizações têm desenvolvido a DMPTool para ajudar os pesquisadores a atender a esses novos requisitos.

Gestão de dados coletados ou gerados na atividade de pesquisa é o foco atual do movimento da Ciência Aberta mediante a crescente exigência do Plano de Gestão de Dados, elaborado de acordo com áreas científicas. Para fortalecer esse movimento, o Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict) lançou um “Manifesto de Acesso Aberto a Dados da Pesquisa Brasileira”, pelo qual estimula a adoção da Ciência Aberta no Brasil, e reivindica amplo e irrestrito acesso a fontes primárias de pesquisa, utilizadas por pesquisadores e outros segmentos sociais.

Cabe lembrar que a Fundação Oswaldo Cruz (Fiocruz) tem desenvolvido várias ações de apoio à implementação e promoção da abertura de dados da instituição, mediante cursos de capacitação, palestras, entre outros, a fim de intensificar as pesquisas relacionadas à Ciência Aberta e à gestão de dados de pesquisas científicas. Mesmo com todas essas iniciativas, a Fiocruz também não tem formulário próprio para aplicação do PGD.

Conforme visto (Seção 3.1), a definição de dados, e também a maneira como eles são utilizados, é complexa e as ferramentas em uso variam de acordo com as disciplinas. As razões para não se inserir nas práticas de CA nem compartilhar os dados são inúmeras. Entretanto, em decorrência dessa complexidade, algumas instituições estão estabelecendo políticas e planejamentos de novos serviços para apoiar a gestão de dados de pesquisa. Todavia, o avolumamento das produções científicas e o requinte dos modernos sistemas de informação científica necessitam de ferramentas e tratamentos dos dados que serão compartilhados. Por isso, a relevância da criação de um plano de gestão de dados, permitirá ampla disseminação das novas descobertas e invenções da comunidade científica.

3.2.3 Princípios FAIR e repositórios de dados de pesquisa

A discussão até agora mostrou o conceito de dados e dados de pesquisa com ênfase no seu acesso e reproduzibilidade. Embora os dados de pesquisa apresentem oportunidades de inovação e descoberta, há inúmeros desafios para disponibilizá-los abertamente, bem como tratar, processar e analisar esses dados. Preocupados com a qualidade do conjunto de dados apresentados, a *Declaration on access to research data from public funding*, publicada pela OECD, em 2004, resultado da reunião de ministros de Ciência e Tecnologia de vários países

estabeleceu um guia internacional com objetivos e princípios direcionados para o acesso aos dados de pesquisa. Citam-se, dentre eles: transparência, conformidade legal, responsabilidade formal, profissionalismo, proteção da propriedade intelectual, interoperabilidade, qualidade e segurança, eficiência e prestação de contas para a sociedade; posteriormente. Em 2007, acrescentaram-se mais quatro princípios: qualidade, segurança, flexibilidade e sustentabilidade (ORGANIZATION FOR ECONOMIC COOPERATION AND DEVELOPMENT, 2007).

Uma boa gestão de dados é primordial na reutilização por meio da comunidade, facilitando a descoberta de novos conhecimentos. Para uma boa gestão, algumas iniciativas foram tomadas. A primeira foi o Data Seal of Approval (DSA), que avaliou repositórios de dados no período de 2020-2021. Essa instituição foi criada em 2008, pela Data Archiving and Networked Services (DANS), posteriormente substituída pelo CoretrustSeal, que tem por objetivo criar selos de aprovação para dados, ajudando a garantir que os arquivados sejam encontrados e reutilizados. Atendendo aos requisitos exigidos pela instituição, os repositórios de dados são armazenados de forma confiável e podem ser utilizados. No Quadro 5, são apresentados alguns dos requisitos exigidos pela CoretrustSeal para avaliação de repositório de dados.

Quadro 5 - Requisitos exigidos pelo CoretrustSeal em 2020-2022 para avaliação de repositório de dados

Requisito a ser considerado	Exigência
Informações sobre direitos de acesso	É preciso constar as informações que identificam as restrições de acesso relativas ao conteúdo das informações, incluindo o arcabouço legal, termos de licenciamento e controle de acesso. Ele contém o acesso e condições de distribuição declaradas dentro do acordo de submissão, relacionadas tanto à preservação e uso final. Também inclui as especificações para a aplicação de direitos.
Avaliação	O processo de determinar se os registros e outros materiais têm permanente (arquivamento) valor. A avaliação pode ser feita na coleção, criador, série, arquivo ou nível de item. A avaliação pode ocorrer antes da doação e antes da transferência física, em ou após a adesão.
Arquivo	Uma organização, lugar ou coleção que armazena registros digitais e analógicos estáticos, documentos e outros materiais, por muito tempo-preservação, de tal forma que possa ser acessado e utilizado por um Designado Comunidade.
Autenticidade	O grau em que uma pessoa (ou sistema) considera um objeto como o que se pretende ser. A autenticidade é julgada com

	base em evidências.
Consumidor	O papel desempenhado por essas pessoas, ou sistemas de clientes, que interagem com serviços [repositórios] para encontrar informações preservadas de interesse e acessá-las em detalhes. Isso pode incluir outros repositórios, bem como pessoas ou sistemas de repositórios internos.
Curadoria	A atividade de gestão e promoção do uso de dados, a partir de seu ponto de criação, visando garantir que eles são adequados aos propósitos contemporâneos e disponíveis para descoberta e reutilização. Relativo a conjuntos de dados dinâmicos, isso pode significar enriquecimento contínuo ou atualização para mantê-los aptos ao propósito. Níveis mais altos de curadoria também envolvem <i>links</i> com anotação e com outros materiais publicados.
Dados	Uma representação reinterpretabel da informação de maneira formalizada adequada para a comunicação, interpretação ou processamento. Exemplos de dados incluem uma sequência de <i>bits</i> , uma tabela de números, ou a gravação de sons feitos por uma pessoa falando.
Base de dados	Coleta de dados organizada de acordo com uma estrutura/modelo conceitual, descrevendo características desses dados e as relações entre suas entidades correspondentes, apoiando uma ou mais áreas de aplicação. Um banco de dados permite que seu conteúdo seja facilmente acessado, gerenciado e atualizado... Um banco de dados pode abrigar um ou muitos conjuntos de dados.
Dataset	Qualquer coleta organizada de dados em formato computacional, definida por um tema ou categoria que reflete o que está sendo medido/observado/monitorado. A apresentação dos dados no aplicativo é habilitada por intermédio de metadados.
Comunidade Designada	Um grupo identificado de potenciais consumidores, que devem ser capazes de entender um determinado conjunto de informações. A Comunidade Designada pode ser composta por várias comunidades de usuários. Uma Comunidade Designada é definida pelo [repositório] e essa definição pode mudar com o tempo.
Preservação Digital	A série de atividades gerenciadas necessárias para garantir o acesso contínuo aos materiais digitais pelo tempo que for necessário. A preservação digital é definida de forma ampla, e se refere a todas as ações necessárias para manter o acesso a materiais digitais além dos limites de falha de mídia ou mudança tecnológica. Esses materiais podem ser registros criados durante o dia a dia de uma organização; “nascido-digital” materiais criados para fins específicos (por exemplo, recursos didáticos); ou os produtos de projetos de digitalização. Essa definição exclui especificamente o uso potencial da tecnologia digital para preservar os artefatos originais mediante digitalização.
Sistema de Informações de Arquivamento Aberto (OAIS)	Uma organização, que pode ser parte de uma organização maior, de pessoas e sistemas, que aceitou a responsabilidade

	de preservar informações e torná-las disponíveis para uma Comunidade Designada. Ele atende a um conjunto de responsabilidades que permite que um Arquivo OAIS seja distinguido de outros usos do termo <i>archive</i> . O termo “aberto” na OAIS é usado para implicar que essa recomendação [i.e., o Modelo de Referência] e futuras recomendações e padrões relacionados são desenvolvidos em aberto, fóruns, e isso não implica que o acesso ao arquivo seja irrestrito.
Formatos preferidos	Formatos que um repositório pode garantir razoavelmente permanecerão legíveis e utilizáveis. Normalmente, esses são os padrões de fato empregados por uma disciplina.
Repositório Institucional	Repositório servindo uma instituição de pesquisa. Muitas vezes são genéricos, multiassuntos repositórios; por exemplo, dentro de uma universidade.

Fonte: Adaptado de Recker e colaboradores (2019, tradução direta).

A segunda iniciativa foi estabelecida em 2016, pelo “The Future of Research Communication and e-Scholarship” (Force11). Essa comunidade definiu os princípios orientados para a publicação de dados – localizável, acessível, interoperável e reutilizável, conhecido pelo acrônimo FAIR (*Findable, Accessible, Interoperable e Reusable*). Cada faceta que forma o acrônimo está relacionada, mas são independentes umas das outras, podendo ser implementadas em qualquer combinação, na medida em que os provedores de dados e FAIRports evoluem para graus crescentes de FAIR-ness. No relatório apresentado, o grupo Force11 mostra os diferentes desafios para quem pretende trabalhar com dados de pesquisa, que dependerá da complexidade e dos interesses dos atores envolvidos, quer sejam pesquisadores, especialistas em dados, comunidade da ciência de mineração de dados ou outros.

Os princípios FAIR são ligados aos princípios de citações de dados definidos na Joint Declaration Data Citation Principles (JDDCP), do Force11, e fornecem orientações que incrementam a implantação de um conjunto de metadados definidos para uso por mecanismos computacionais automatizados, as máquinas e também para uso humano. Ao adotar esses princípios (Figura 6), o dado se torna completamente localizável, acessível, interoperável e reutilizável.

Figura 6 - Princípios FAIR



Fonte: FAIR data training

Segundo Force11 ([2016, on-line, tradução direta]⁴², para ser localizável, qualquer objeto de dados deve ser único, persistente e identificável:

- a) o mesmo objeto de dados deve ser localizável a qualquer momento, portanto, deve ser persistente, com ênfase em seus metadados. Para isso, eles devem ser indexados em mecanismos de busca, que possibilitem aos computadores e usuários encontrá-los com facilidade;
- b) um objeto de dados deve conter minimamente metadados açãoáveis básicos da máquina, permitindo que ele seja distinguido de outros objetos de dados. Ou seja, o conjunto de dados deve ser descrito por metadados ricos o suficiente para que, uma vez indexados em um mecanismo de busca, sejam encontrados mesmo sem o seu identificador único persistente;
- c) os identificadores de qualquer conceito utilizado em objeto de dados devem, portanto, ser únicos e persistentes. Para isso, deve-se adotar um identificador único persistente, tanto para o conjunto de dados quanto para os metadados, por exemplo: DOI.

⁴² A descrição dessas informações encontra-se no site da Force11. Disponível em: <https://www.force11.org/fairprinciples>. Acesso em: 1 set. 2021.

Os dados são acessíveis quando podem ser sempre obtidos por máquinas e seres humanos, isso quando for:

- a) mediante a devida autorização: a depender das restrições de acesso aos dados e/ou metadados, um mecanismo de autenticação e autorização para o acesso deve ser liberado pelo protocolo de comunicação; como exemplo, os repositórios confiáveis oferecem essa opção;
- b) por meio de um protocolo bem definido: com o identificador persistente do conjunto de dados e/ou de seus metadados, o usuário deverá recuperá-los mais facilmente por meio de protocolos de comunicação padronizados; por exemplo, HTTP ou FTP;
- c) assim, tanto as máquinas quanto os seres humanos poderão julgar a real capacidade de acesso de cada objeto de dados. Independente de licenciamento dos dados e dos metadados, o protocolo de comunicação usado para acessá-los deve ser aberto, gratuito e passível de ser implementado por qualquer interessado; por exemplo HTTP ou FTP.

A interoperabilidade dos dados acontece se:

- a) os metadados são acionáveis por máquina. Nesse caso, devem ser representados por meio de uma linguagem formal, linguagens de representação do conhecimento que sejam padronizadas, acessíveis e amplamente aplicáveis; por exemplo: RDF, XML, DICOM, entre outros;
- b) os metadados usam os vocabulários compartilhados e/ou ontologias compartilhadas que seguem os princípios FAIR. Nesse item, os dados e metadados devem possuir referências a vocabulários e/ou ontologias que os descrevem, além de garantir que também sigam os princípios FAIR;
- c) os metadados dentro do objeto de dados devem ser tanto sintaticamente interpretados como semanticamente acessíveis para máquina.

Para que o objeto de dados seja reutilizável, os critérios adicionais são:

- a) os dados devem estar em conformidade com os princípios anteriores;
- b) os metadados devem ser suficientemente bem descritos e ricos para que possam ser automaticamente (ou com o mínimo esforço humano) vinculados ou integrados – *like-with-like*, – com outras fontes de dados;
- c) os dados publicados devem se referir às suas fontes com metadados ricos e

procedência suficiente para permitir a citação adequada.

Os princípios FAIR servem como apoio na gestão de dados e divulgação de pesquisas; além disso, contribuem para a qualidade do registro de pesquisa, pois ajudam os pesquisadores a organizar e tratar os dados. Diante da dificuldade de implementar os princípios FAIR, foi criado, pelos governos da Alemanha, Holanda e França, a iniciativa GO FAIR com o intuito orientar o processo de implantação a ser utilizado no desenvolvimento da interoperabilidade de dados. Como resultado dessa iniciativa, estabeleceram-se ações que foram desenvolvidas na Holanda, em 2020: promoção do acesso às publicações científicas (*open access*); promoção da utilização e reutilização dos dados de pesquisa (*open data*); e adaptação aos sistemas de avaliação e de propriedade intelectual, todos direcionados para o fortalecimento da Ciência Aberta.

A Figura 7 mostra a iniciativa GO FAIR, que se baseia em três pilares: Go-Change, Go-Train e Go-Build. Ela tem por prioridade tornar mais facilmente localizável, acessível, interoperável e reutilizável todos os tipos de dados que atualmente se encontram subdivididos e desconectados.

Figura 7 - Atividades da GO FAIR



Fonte: go-fair.org

Go-Change tem essa iniciativa em focar nas políticas e prioridades, proporcionando mudanças socioculturais que envolvem relevantes atores de todos os níveis, a fim de tornar os princípios FAIR um padrão de trabalho na ciência; além disso, restaurar os sistemas de recompensa, de modo que a avaliação das carreiras de pesquisa reconheça as atividades de

ciência aberta.

Go-Train trata da definição de currículos e programas de treinamento a respeito do gerenciamento de dados. Para isso, promove treinamento de capacitação de profissionais em gestão de dados.

Go-Build traz como foco o tema das infraestruturas para dados interoperáveis, criando padrões, protocolos e serviços compatíveis e permitindo que os pesquisadores depositem, acessem e analisem dados científicos de todas as áreas.

No Brasil, as iniciativas do GO FAIR chegaram por meio do Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict), e sua responsabilidade é difundir, apoiar e coordenar as atividades relacionadas à adoção das estratégias de implementação dos princípios FAIR, de acordo com as especificidades das áreas do conhecimento, em todo o país.

Na perspectiva dos princípios FAIR, a gestão de dados de pesquisa é um instrumento facilitador durante todo o ciclo de vida da pesquisa, pois estabelece orientações substanciais e um guia na gestão, curadoria e preservação dos dados de pesquisas, com a finalidade de compartilhar e reutilizar. Assim, ao se construir um Plano de Gestão de Dados, faz-se necessário um alinhamento com os princípios FAIR.

Para isso, é preciso oferecer um ambiente dinâmico e flexível para dar suporte na execução dos processos e dos repositórios de dados de pesquisa, pela natureza dos dados, desempenham essa função. Tal apoio requer um arcabouço tecnológico e gerencial que compreenda o ciclo de vida dos dados, pois, ao se construir um projeto de pesquisa, o pesquisador precisa pensar onde serão armazenados e preservados os dados de suas pesquisas. No âmago desse arcabouço estão os repositórios de dados, visto que se tornam parte fundamental da infraestrutura mundial de pesquisa e cumprem seu papel de organizar, armazenar e possibilitar o acesso e o uso de resultados apresentados em diferentes formatos.

Os repositórios de dados de pesquisa⁴³ são as ferramentas necessárias para armazenar e gerenciar os dados utilizados e produzidos no seu percurso. Ao definir repositório de dados de pesquisa, Uzwyshyn (2016, tradução nossa) assevera que: “repositórios on-line de dados de pesquisa são bases de dados de grande estrutura que tem sua configuração focada na gerência, compartilhamento, acesso e arquivamento de dados de pesquisadores”. Da mesma

⁴³ Para obter informações sobre repositórios públicos de dados, vale a pena consultar o *R3Data – Registry of Research Data Repositories* – uma ferramenta de busca de repositórios de dados.

forma, Monteiro (2017) afirma que os repositórios de dados de pesquisa são uma evolução dos repositórios documentais, e são ambientes instituídos em universidades para apoiar os pesquisadores no processo de gerenciamento e disponibilização dos dados de pesquisa. Para o autor, irá contribuir, dessa maneira, para que outros pesquisadores possam reutilizá-los e auxiliar no gerenciamento de dados. Isso é fundamental que sejam planejados e estruturados desde o início de sua implementação.

Além da institucional, os repositórios são caracterizados, com algumas sobreposições e diferenças, como institucionais, disciplinares, multidisciplinares e orientados por projeto (SAYÃO; SALES, 2016; SILVA, 2019):

- a) repositórios institucionais de dados de pesquisa são os que têm o papel de coletar, administrar e armazenar documentos completos de produções intelectuais de uma instituição acadêmica;
- b) repositórios disciplinares ou temáticos são os que apresentam dados voltados para as disciplinas, por exemplo, o *Database* – repositório voltado para arquivamento, descoberta e intercâmbio de modelos computacionais na área de biologia;
- c) repositórios de dados multidisciplinares são os destinados a arquivar dados resultantes de pesquisas múltiplas áreas de conhecimento (muitos repositórios institucionais também podem ser denominados como multidisciplinares, devido à natureza das instituições);
- d) repositórios de dados de projetos são gerados a partir dos resultados de pesquisas destinadas à solução de problemas específicos, e estão bastante relacionados a órgãos governamentais e não governamentais e empresas privadas;
- e) repositórios de dados de propósito geral são aqueles que qualquer pesquisador pode usar, tendo ou não filiação institucional; por exemplo, o Figshare e Zenodo;
- f) repositórios de dados editoriais são similares aos institucionais, porém apresentam características especiais para comunidade específicas, exemplo o Dryad⁴⁴;

⁴⁴ Cabe lembrar que os repositórios Dataverse, DataHub, DANS, e os Re却itórios Multidisciplinares como Figshare, Dryad e Zenodo, seguem os Princípios FAIR (*Findable, Accessible, Interoperable, Reusable*).

g) repositórios próprios, que são os lugares onde os pesquisadores armazenam seus dados de pesquisa, localizados em nuvem. Podem ser um servidor pessoal ou de projetos; por exemplo: Amazon Cloud Drive.

As universidades federais têm um importante papel na disponibilização de recursos e laboratórios abertos, e assim tornam-se grandes incentivadoras ao acesso aberto. Tanto a Ciência Aberta quanto a implementação do Plano de Gestão de Dados de Pesquisa nas universidades são fundamentais para a democratização da Ciência, pois entende-se que a divulgação e o acesso aos dados de pesquisa científica é um dever social.

4 PERCURSO METODOLÓGICO

Para atingir os objetivos propostos, serão apresentados, neste capítulo, os procedimentos metodológicos que nortearam o desenvolvimento desta pesquisa, agregando não apenas a teoria, mas também a classificação da pesquisa, o detalhamento do campo empírico, o universo no qual a pesquisa será contextualizada e a população-alvo. Ademais, apresentar-se-á os instrumentos de coleta de dados, bem como a análise, a fim de obter melhor detalhamento para a interpretação dos resultados.

Com base no referencial teórico, percebeu-se que as transformações ocorridas no contexto da ciência mudaram significativamente a maneira como os pesquisadores realizavam suas atividades. Tal mudança ocorreu, principalmente, devido à inserção de novos equipamentos e instrumentos mais requintados para uso, quer seja no percurso ou após a conclusão da pesquisa. Com o objetivo de compreender essas evoluções paradigmáticas tradicionais da ciência e suas alterações nos modos de produção – bem como no uso das tecnologias para esse fim e a contribuição das universidades no processo de comunicação científica, gestão de dados da pesquisa como elemento imprescindível na promoção da CA –, a abordagem teórica foi fundamentada na literatura internacional e nacional, publicada a respeito de gestão de dados de pesquisa e os assuntos adjacentes, com a finalidade de trazer reflexões filosóficas e científicas. Filosóficas não apenas com o intuito de apresentar novas realidades, mas visando revelar aquilo que já se sabe, porém não se sabe exatamente como; e científicas, pois pretende-se evidenciar o desconhecido.

Em conformidade com os tipos de pesquisa apresentados na literatura e destacados pelos autores Piovesan e Temporini (1995), Vergara (2003), Quivy e Campenhoudt (2005), Yin (2005) e Bardin (2011), o percurso metodológico aplicado neste estudo segue os critérios e as normas identificados na literatura, empregando método e procedimentos específicos para o progresso de cada etapa. Para tanto, as etapas abordadas são as seguintes: a) delineamento da pesquisa; b) definição constitutiva e operacional das variáveis; c) elaboração do instrumento e procedimentos de coleta de dados; d) identificação do universo; e) estudo de caso; e, f) análise dos dados.

4.1 ABORDAGEM METODOLÓGICA

A presente pesquisa foi norteada com o seguinte questionamento: como a Universidade Federal da Bahia tem desenvolvido as ações, no contexto da Ciência Aberta, para a implantação de políticas de gestão de dados de pesquisa, e quais as percepções e os

desafios dos professores nesse processo? Com base nos objetivos e no problema estabelecido que constam na apresentação deste projeto, evidenciou-se a necessidade de se aprofundar no sujeito da pesquisa, por se tratar de um fenômeno contemporâneo, histórico e, no caso, único, remetendo-se a um estudo de caso.

Sendo assim, tornam-se relevantes as singularidades dos sujeitos e tema estudado a fim de compor a análise a respeito da realidade do problema. Nesse sentido, Creswell (2015) apresenta três fatores envolvidos no projeto de pesquisa: a concepção filosófica; a estratégia de investigação, que se relaciona com a concepção escolhida; e os métodos ou procedimentos de pesquisa específicos, que modificam a abordagem em prática. Para o autor, a pesquisa se inicia pelos pressupostos filosóficos, principalmente os de natureza qualitativa; mesmo estando em fase oculta, essas concepções são norteadoras da ação.

Dentre as concepções filosóficas apresentadas por Creswell (2015), e de acordo com a sistematização apresentada pelo autor, a pragmática foi escolhida como método neste estudo por ter como base a concentração nos problemas da pesquisa e não nos métodos. Além disso, a concepção filosófica pragmática utiliza todas as abordagens disponíveis para resolver os problemas; ou seja, o pesquisador possui mais liberdade de escolha dos métodos, técnicas e procedimentos que melhor se adequem às necessidades e finalidade do estudo. Uma vez conhecendo o método, definem-se as técnicas e os procedimentos.

Trata-se de uma pesquisa empírica, tendo como campo de estudo a UFBA, especificamente os pesquisadores/docentes dos cursos de pós-graduação *stricto sensu* (mestrado e doutorado) que compõem as cinco áreas de conhecimento; ou seja, Ciências Exatas, Biológicas, Ciências Humanas, Letras e Artes. Essa modalidade configura-se oportunamente, pois busca-se conhecer e averiguar a prática da gestão de dados de pesquisa em um universo acadêmico. A apreciação desse tipo de pesquisa é pela viabilidade que oferece na solidez das argumentações, por mais espessa que seja a base fatual. É um estudo embasado na experiência, doutrina ou atitude do fenômeno a ser estudado, enriquecendo a fundamentação dos argumentos de pesquisa.

Uma pesquisa é classificada em três grandes grupos: exploratória, descritiva e explicativa (PIOVESAN; TEMPORINI, 1995). Por se tratar de uma pesquisa empírica, pela qual se pretende proporcionar maior familiaridade com o problema, tornando-o mais explícito, e envolver procedimentos como levantamento bibliográfico, pesquisa bibliográfica, entrevistas com pessoas relacionadas à pesquisa e análise de exemplos, adotou-se a concepção

descritiva. Quanto aos propósitos apresentados, esta pesquisa é considerada descritiva, já que dela será necessário obter maior fundamento, clareza e completude com a descrição efetiva dos elementos coletados e analisados, para descobrir a existência de possíveis associações entre as variáveis, que, no caso desta pesquisa, serão: os resultados da pesquisa bibliográfica, as suas relações com o conjunto de dados coletados por meio dos instrumentos e a análise obtida mediante a unidade de análise (“o caso”). Para Vergara (2003), um estudo de natureza descritiva busca descrever, de maneira detalhada, os fenômenos de determinada realidade e população. Dessarte, este estudo é exploratório-descritivo. Predominantemente exploratório porque carece de dados provenientes de pesquisas científicas anteriores, e descritivo porque objetiva conhecer e descrever o contexto investigado. E isso se aplica ao lócus do nosso estudo, ou seja, a UFBA.

Como o assunto abordado nesta pesquisa traz questões que envolvem aspectos colaterais complexos – por exemplo, propriedade intelectual, direito autoral, autoria colaborativa e comunicação por meio de tecnologias, no que se refere à natureza –, será apropriado utilizar a pesquisa aplicada, pois pretende-se gerar conhecimentos para aplicação prática, visando a solução de problemas; ou seja, envolve verdades e interesses locais. Michel (2009) ainda argumenta que esse tipo de pesquisa se baseia na coleta de fenômenos que ocorrem na realidade a ser pesquisada, e é justamente a proposta deste trabalho.

Após definidos o tipo de pesquisa (empírica), os objetivos (descritivo) e a natureza (aplicada), é importante entender e definir a maneira que os dados serão coletados (YIN, 2005). Dessa forma, será adotado o estudo de caso como estratégia de pesquisa, pois se pretende observar e conhecer a concepção dos pesquisadores quanto ao tema em questão (gestão de dados de pesquisa) dentro de um lugar específico (UFBA). Para tanto, inicia-se com a revisão minuciosa da literatura e a proposição cuidadosa e atenta das questões e dos objetivos da pesquisa.

Em função do objeto de estudo e vigilante às novas perspectivas de compartilhamento, acesso e uso dos dados de pesquisas, a pesquisadora, por residir no estado, saber tratar-se de uma universidade pública, reconhecida pelo seu pioneirismo nos estudos interdisciplinares na Bahia, e além de atenta aos fenômenos digitais de tratamento, recuperação e preservação da informação – ao mesmo tempo ansiosa com o arcabouço que envolve uma política de informação –, selecionou a Universidade Federal da Bahia para ser o lócus e os professores pesquisadores como objeto de estudo para esta pesquisa.

No que tange à abordagem, será qualitativa. Qualitativa porque busca-se “captar o fenômeno em estudo a partir da perspectiva das pessoas nele envolvidas, considerando todos os pontos de vista relevantes [...]” (GODOY, 1995, p. 21). Para o autor, a abordagem qualitativa “[...] não se apresenta como uma proposta rigidamente estruturada, ela permite que a imaginação e a criatividade levem os investigadores a propor trabalhos que explorem novos enfoques” (GODOY, 1995, p. 23), não requer o uso de método e técnicas estatísticas, pois permite o mundo objetivo e subjetivo dos sujeitos.

Diferentemente das ciências naturais, nas ciências sociais, em geral, os fenômenos são complexos, pois não podem ser reproduzidos em laboratório e submetidos a controle; ou seja, as interpretações dos fenômenos dos métodos qualitativos são analisadas indutivamente (FERNANDES; GOMES, 2003). Logo, diante da diversidade de perspectiva, o fazer científico não segue um único padrão de trabalho; pelo contrário, a sociologia sempre foi marcada pela variedade de técnicas/métodos, quer sejam de investigação ou de explicação. Diante disso, o método qualitativo tem por características a flexibilidade – principalmente no que se refere às técnicas de coletas de dados –, e a heterodoxia no momento da sua análise. Isso exige do pesquisador uma capacidade integrativa e analítica que, de certo modo, depende do desenvolvimento de uma capacidade criadora e intuitiva (MARTINS, 2004).

Ao contrário da abordagem qualitativa, a quantitativa considera que tudo pode ser quantificável, o que significa traduzir, em números, opiniões e informações para, posteriormente, classificar e analisar cada dado usando recursos e técnicas estatísticas (MICHEL, 2009). O uso de uma metodologia em detrimento da outra não significa dizer que uma seja superior a outra; a escolha depende do tipo de problema a ser pesquisado e dos objetivos da pesquisa.

4.2 UNIVERSO, AMOSTRA E POPULAÇÃO-ALVO DA PESQUISA

Esta seção apresenta o universo e a população-alvo deste estudo. Assim, ao abordar sobre a população ou universo da pesquisa, Vergara (2011) afirma se tratar de um conjunto de elementos que possuem características que serão objeto de estudo; ou seja, o total de indivíduos a serem analisados. Observa ainda Oliveira (2005) que nem sempre é possível pesquisar o universo devido aos recursos necessários e, muitas vezes, falta tempo para o levantamento de dados desse tipo; por conseguinte, estuda-se a parte desse grupo, chamada amostra. Amostra, na concepção de Vergara (2011), é a parte do universo selecionada a partir de um critério de representatividade, e com isso o levantamento por amostragem apresenta

menor custo, menos tempo, objetivos mais amplos e dados mais fidedignos; o público-alvo, portanto, é aquele que se tem interesse em estudar.

No entanto, ao aplicar o teste piloto, percebeu-se pouca adesão no retorno do instrumento de coleta de dados, o que poderia comprometer a análise do estudo. Nesse sentido, optou-se em aplicar o questionário para toda a população, ou seja, 931 professores do quadro permanentes da pós-graduação da UFBA, dispensando o uso da amostragem porque não seria representativa neste estudo. Sendo assim, o público-alvo, portanto, é aquele que se tem interesse em estudar.

Cabe enfatizar que, em uma pesquisa que utiliza o método estudo de caso, o universo se confunde com a amostra. Por isso, compete ao pesquisador fixá-la para não gerar dúvidas no leitor. Quanto a esta pesquisa, o universo/população refere-se a todos os professores do quadro permanente dos programas de pós-graduação (mestrado e doutorado acadêmico - M/D) da UFBA que obtiveram conceito Capes igual ou superior a 4 no quadriênio 2017-2021; excluindo, portanto, o mestrado profissional (MP). A população-alvo foi os professores pesquisadores do quadro permanente que compunham o programa de pós-graduação acadêmico selecionado pelo critério previamente estabelecido.

A UFBA possui 97 cursos de graduação (incluindo diurno e noturno, bacharel e licenciatura), organizados em cinco áreas: Área I - Ciência Exatas, Área II - Ciências Biológicas/Saúde, Área III - Ciências Humanas, Área IV - Letras e a Área V - Artes. Além dessas, oferece sete cursos de graduação na modalidade a distância (sendo que cinco estão em andamento e dois encerrados), a área Bacharelado Interdisciplinar, com oito cursos (noturno e diurno), e dois cursos em tecnologia da área, pertencente ao Curso Superior em Tecnologia⁴⁵.

Em relação aos cursos de pós-graduação, a UFBA possui 83 que estão distribuídos em cinco áreas, sendo: Área I - 23 cursos, Área II - 24 cursos, Área III - 26 cursos, Área IV - 3 cursos e Área V - 7 cursos. No Quadro 6, são apresentados o total dos cursos de pós-graduação, por área de conhecimento, e o quantitativo de cursos que obtiveram conceito igual ou maior que quatro (≥ 4) no quadriênio 2017-2021.

⁴⁵ Para mais informações ver: <https://heyzine.com/flip-book, 2022.>

Quadro 6 - Cursos de Pós-Graduação da UFBA com respectivo conceito Capes no quadriênio 2017-2021

ÁREA	#	CURSO	NÍVEL (Mestrado - M; Doutorado - D; Mestrado Profissional - MP)	CONCEITO CAPES (M/D)
I - Ciências Exatas	1	Arquitetura e Urbanismo	M/D	6
	2	Ciência da Computação	M/D	5
	3	Ciência da Computação (UFBA/UEFS/UNIFACS)	D	5
	4	Conservação e Restauração de Monumentos e Núcleos Históricos	MP	5
	5	Energia e Ambiente	D	4
	6	Engenharia Civil	M/D	5
	7	Engenharia Elétrica	M/D	4
	8	Engenharia Industrial	M/D	5
	9	Engenharia Industrial	MP	4
	10	Engenharia Química	M	4
	11	Engenharia Química UFBA/UNIFACS	D	3
	12	Física	M/D	4
	13	Geofísica	M/D	4
	14	Geografia	M/D	4
	15	Geologia	M/D	4
	16	Geoquímica: Petróleo e Meio Ambiente	M/D	4
	17	Gestão e Regulação de Recursos Hídricos	M	-
	18	Matemática	M/D	4
	19	Matemática em Rede (Profmat)	MP	5
	20	Mecatrônica	M/D	4
	21	Meio Ambiente, Águas e Saneamento	M	4
	22	Propriedade Intelectual e Transferência de Tecnologia para Inovação	MP	4
	23	Química	M/D	4
	1	Alimentos, Nutrição e Saúde	M/D	4
	2	Assistência Farmacêutica	M/D	4

II - Ciências Biológicas/Saúde	3	Biociências	M	3
	4	Biodiversidade e Evolução	M/D	5
	5	Biotecnologia	M/D	5
	6	Biotecnologia (Renorbio)	D	6
	7	Ciências da Reabilitação	M	-
	8	Ecologia Aplicada à Gestão Ambiental	MP	4
	9	Ecologia: Teoria, Aplicada e Valores	M/D	6
	10	Enfermagem e Saúde	M/D	5
	11	Farmácia	M/D	4
	12	Imunologia	M/D	5
	13	Medicina e Saúde	M/D	5
	14	Microbiologia	M	3
	15	Multicêntrico em Bioquímica e Biologia Molecular (SBBq)	M/D	5
	16	Multicêntrico em Ciências Fisiológicas	M/D	5
	17	Odontologia e Saúde	M/D	5
	18	Patologia Humana	M/D	6
	19	Processos Interativos dos Órgãos e Sistemas	M/D	5
	20	Saúde Coletiva	MP	5
	21	Saúde Coletiva (IMS)	M	4
	22	Saúde Coletiva (ISC)	D	7
	23	Saúde, Ambiente e Trabalho	M	4
	24	Zootecnia	M/D	4
III - Ciências Humanas	1	Administração	M/D	5
	2	Administração	MP	4
	3	Antropologia	M/D	4
	4	Ciências da Informação	M/D	4
	5	Ciências Sociais	M/D	4
	6	Ciências Política	M	-
	7	Comunicação e Cultura Contemporâneas	M/D	6
	8	Contabilidade	M	4
	9	Curriculum, Linguagens e Inovações Pedagógicas	MP	4

	10	Desenvolvimento e Gestão Social	MP	3
	11	Difusão do Conhecimento	D	5
	12	Direito	M/D	5
	13	Economia	M/D	4
	14	Educação	M/D	5
	15	Ensino, Filosofia e História da Ciências	M/D	5
	16	Estudo Étnicos e Africanos	M/D	4
	17	Estudos Interdisciplinares sobre a Universidade	M	4
	18	Estudos Interdisciplinares sobre Mulheres, Gênero e Feminismo	M/D	4
	19	Filosofia	M/D	5
	20	História	M/D	5
	21	Museologia	M	3
	22	Psicologia	M/D	6
	23	Psicologia da Saúde	MP	-
	24	Relações Internacionais	M	4
	25	Segurança Pública, Justiça e Cidadania	MP	3
	26	Serviço Social	M	-
IV - Letras	1	Letras	MP	4
	2	Língua e Cultura	M/D	5
	3	Literatura e Cultura	M/D	5
V - Artes	1	Artes cênicas	M/D	6
	2	Artes (Profartes)	MP	4
	3	Artes visuais	M/D	4
	4	Dança	M/D	5
	5	Dança	MP	-
	6	Música	MP	4
	7	Música	M/D	4

Fonte: <https://heyzine.com/flip-book>, 2022.

Com base no exposto, a população está exposta na Tabela 1; na qual apresenta o total dos cursos de pós-graduação, por área de conhecimento, o quantitativo de cursos de

pós-graduação (M/D) que obtiveram conceito igual ou maior que quatro (≥ 4), exceto os cursos de mestrado profissional (MP), e o número de professores do quadro permanente. Assim, do total de 83 cursos, 61 (73,49%) cursos contém conceito igual ou maior que quatro (≥ 4), contudo, apenas 39 cursos continham os dados completos a respeito de quem seriam os professores permanentes e os e-mails de cada um, em seus respectivos *sites*. Desta maneira, e com base no critério previamente estabelecido, a população é equivalente a 931 professores permanentes dos cursos de pós-graduação que constavam dados completos em seus sites (identificação de quem seriam os professores permanentes e e-mails), conforme a tabela abaixo. Destes, 51 professores se encontram em dois ou três programas de pós-graduação/UFBA, logo o instrumento foi aplicado a 880.

Tabela 1 – População da pesquisa

Área de conhecimento	Total de cursos	Cursos com conceito ≥ 4	Cursos com dados completos	Professores permanentes
Ciências Exatas	23	17 (73,91%)	12	264
Ciências Humanas	26	18 (69,23%)	11	230
Ciências Biológicas	24	20 (83,33%)	11	289
Letras	3	2 (66,67%)	2	84
Artes	7	4 (57,14%)	3	64
Total	83	61 (73,49%)	39	931

Fonte: Elaboração própria.

Os nomes dos professores permanentes, bem como os respectivos e-mails, foram coletados pelo *site* e/ou por solicitação à Secretaria dos respectivos programas, previamente selecionados. Após tentativas de obter as informações corretas na coordenação dos respectivos cursos, sem êxito, optou-se em excluí-los em decorrência da falta de informações que pudessem distinguir professor permanente, de colaborador e de visitante, bem como da falta de contatos para envio do questionário. Essas tentativas foram feitas via e-mail e telefônico para a secretaria e também para as coordenações dos cursos que não obtinham as informações completas (e-mail e/ou distinção do quadro de professor). Diante disso, a população estudada foi de 880 professores permanentes.

Após essas etapas, aplicou-se o questionário aos professores pertencentes a população já que o quantitativo representaria substancialmente a proposta deste estudo, contendo o nível

de significância de 95%.

4.3 PROCEDIMENTOS

A fim de atender aos objetivos propostos, descrever a unidade social – que, no caso é a Universidade Federal da Bahia – e compreender a temática, realizou-se um levantamento em trabalhos existentes na área, livros, monografias, dissertações e teses, além de artigos publicados em revistas técnicas especializadas e eventos da área da Ciência da Informação. O levantamento bibliográfico possibilita captar informações essenciais no embasamento teórico, para fundamentar a pesquisa. Dessa maneira, utilizaram-se as seguintes bases de dados: a Scientific Electronic Library Online (Scielo) – resultado de um projeto de pesquisa da Fundação de Amparo à Pesquisa do Estado de São Paulo (Fapesp) e Biblioteca Digital Brasileira de Teses e Dissertações (BDTD), desenvolvida e coordenada pelo Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict); Base de Dados Referenciais de Artigos de Periódicos em Ciência da Informação (Brapci); bem como pesquisas apresentadas em eventos da área de Ciência da Informação. Sobre isso, Vergara (2011, p. 43) afirma o seguinte:

[...] quanto aos meios, uma pesquisa é classificada como bibliográfica, quando a revisão de literatura é baseada em fontes públicas nacionais e internacionais, tais como portal de periódicos, artigos publicados em anais de congressos, livros, dissertações e teses, com o objetivo de fornecer subsídios para um referencial teórico embasado sobre o assunto a ser discutido.

Nesse sentido, algumas listas de referências dos trabalhos pesquisados foram fontes fundamentais, fornecendo dados e informações necessárias para a execução da pesquisa. Sabe-se que uma pesquisa engloba levantamento bibliográfico, pesquisa bibliográfica, narrativa em diários, questionários, entrevistas com pessoas que tiveram ou têm experiências com o fenômeno estudado, pesquisa bibliográfica e outros elementos que permitam a compreensão e a interpretação dos dados (QUIVY; CAMPENHOUDT, 2005).

Para obter outras descobertas que possam acrescentar aspectos novos à problemática, será desenvolvido um estudo de caso.

4.3.1 Estudo de caso

Por se tratar de uma pesquisa descritiva, o estudo de caso aprofundará e ampliará a realidade pesquisada, no caso, a UFBA. Embora existam outras instituições universitárias que já apresentam trabalhos semelhantes a respeito da implantação do plano de gestão de dados de

pesquisa – como da Universidade de São Paulo (USP)⁴⁶ –, a Universidade Federal da Bahia foi escolhida por ser pública, compor um dos primeiros cursos universitários no Brasil – a Escola de Cirurgia da Bahia, hoje Faculdade de Medicina da Bahia – e pela viabilidade no acesso às informações. Sendo o acesso a dados de pesquisas tão importante para a valorização da ciência, busca-se estudar a inserção desse movimento nesse lócus universitário.

A missão da universidade é baseada em ensino, pesquisa e extensão, e construída de forma colaborativa. Em virtude da Ciência Aberta, ocorreu uma mudança de pesquisadores individuais trabalhando de forma independente visando a formação de grandes institutos, onde as pessoas se reúnem para pesquisar conjuntamente. Ultimamente, há maior consciência entre os pesquisadores de que a tecnologia da informação se tornou um componente indispensável na realização das pesquisas, no gerenciamento de informação e dados e, sobretudo, na tomada de decisão. Para fortalecer as inovações tecnológicas na abertura e compartilhamento dos dados, bem como o avanço da ciência, as universidades – principalmente as internacionais e algumas no Brasil (a passos lentos) – estão se inserindo no movimento da CA. Porém, ainda há algumas barreiras para gerenciar com eficiência e eficácia os dados produzidos nas instituições científicas.

Em decorrência disso, o estudo de caso na UFBA se faz necessário, a fim de atender aos objetivos propostos. Martins (1994) assevera que o estudo de caso se trata de uma constatação baseada na experiência, em que se pesquisa um fenômeno recente dentro do argumento da vida real, quando o limite entre o fenômeno e o contexto não é claramente presente – no caso, o plano de gestão de dados de pesquisas científicas. O estudo de caso, como método de pesquisa, surge do desejo de compreender os fenômenos sociais complexos; ou seja, “[...] permite que os investigadores foquem um ‘caso’ e retenha uma perspectiva holística e do mundo real [...]” (YIN, 2005, p. 4, grifo do autor). Segundo Yin (2005), o estudo de caso, embora tenha pontos comuns com o método histórico, caracteriza-se pela capacidade de lidar com uma completa variedade de evidências, que são obtidas por meio de documentos, artefatos, entrevistas e observações.

Considerando que existe uma relação dinâmica entre o mundo real e o sujeito e que, em uma pesquisa qualitativa, a interpretação dos fenômenos e a atribuição de significados fazem parte do processo básico do seu desenvolvimento, esta tese conduz para a construção de um estudo de caso porque, além de interpretar o contexto de um ambiente específico – no

⁴⁶ Ver: Documentos e páginas de interesse - Planos de Gestão de Dados (fapesp.br).

caso a gestão de dados de pesquisa científica na UFBA –, pretende-se identificar os fenômenos que ali ocorrem (YIN, 2005). Yin (2005, xxxii) explica, também, que a pesquisa de estudo de caso seria o método preferencial quando comparado a outros em situações nas quais o foco de estudo é um fenômeno contemporâneo em seu contexto no mundo real, ao invés de um fenômeno completamente histórico. Sendo a Ciência Aberta o quarto paradigma científico, fenômeno contemporâneo, justifica-se a escolha desse método.

Na concepção de Yin (2005), estudo de caso pode ser compreendido em duas partes: a primeira inicia com o escopo do estudo de caso, ou seja, é uma investigação empírica que estuda o fenômeno contemporâneo em seu contexto do mundo real. Isso acontece quando os limites entre o fenômeno e o contexto não forem evidentes; a segunda emerge quando o fenômeno e o contexto não se apresentam claramente distinguíveis nas situações do mundo real, e com isso outras características metodológicas tornam-se relevantes. Em suma, a definição descrita pelo autor mostra como a pesquisa de estudo de caso compreende um método abrangente e abarca diferentes orientações epistemológicas.

Ao realizar esse tipo de estudo, objetiva-se expandir e generalizar teorias sem alongar o tempo, pois isso confunde estudo de caso com outros métodos específicos, por exemplo, a etnografia (YIN, 2015). As etapas para a realização de um estudo de caso são:

- a) elaboração do projeto de pesquisa com a definição dos seus componentes (questão de pesquisa, proposições teóricas, unidade de análise, vinculação dos dados às proposições, critérios para interpretar as constatações). Nessa etapa, definiu-se a proposição do estudo e os procedimentos para a coleta de dados a partir da questão de pesquisa;
- b) elaboração do protocolo (Figura 8) para o estudo de caso. Nessa etapa, pretende-se elaborar o protocolo para a coleta dos dados, cujo objetivo é apresentar os passos lógicos para a construção do estudo de caso;
- c) coleta de dados. Nessa etapa, pretende-se coletar os dados por meio dos respectivos instrumentos– questionário e entrevista. Para o questionário, o período de coleta dependerá do calendário acadêmico da UFBA. Planeja-se aplicar um mês após o início do período letivo. A entrevista será realizada após a análise dos dados do questionário, pois este é que dará subsídio para o planejamento do roteiro semiestruturado;
- d) relatório individual dos casos, análise dos dados e relatório final com

cruzamento dos dados. Realizar-se-á a análise dos dados por meio de algumas estratégias, técnicas e ferramentas.

O estudo de caso, segundo Yin (2005), se aplica quando se pretende analisar uma unidade profundamente, objetivando o exame detalhado de um ambiente, que pode ser um evento, entidade, ou um único indivíduo. Para o autor, a definição do caso (unidade de análise) está relacionada com a forma de se definir a questão inicial da pesquisa.

Por considerar que os estudos a respeito de plano de gestão de dados de pesquisa científica ainda se encontram em estágios embrionários, no Brasil, e saber da importância histórica que é a Universidade Federal da Bahia, apontada acima, a escolha do método do estudo de caso pareceu ser o mais apropriado para a realização desta pesquisa.

Yin (2005, p. 53) apresenta quatro tipos básicos de estudo de caso:

- a) Projetos de caso único holístico – unidade única de análise e único caso;
- b) Projetos de caso único integrado – unidades múltiplas de análise e único caso;
- c) Projetos de casos múltiplos holísticos – unidade única de análise e múltiplos casos;
- d) Projetos de casos múltiplos integrados – unidades múltiplas de análise e múltiplos casos;
- e) Projetos de métodos mistos – combinação de técnicas e métodos.

A escolha entre os tipos de projetos dependerá da questão da pesquisa. Para isso, é recomendado o projeto holístico quando “não é possível identificar nenhuma subunidade lógica e quando a teoria em questão subjacente do estudo de caso é ela própria de natureza holística”. Por sua vez, o projeto integrado é utilizado quando o estudo de caso, único ou múltiplos casos, envolve subunidades de análise, como exemplo, o pesquisador decide incluir os funcionários como uma subunidade de estudo (YIN, 2005, p. 58).

Sendo assim, a presente pesquisa pode ser definida como estudo de caso único integrado, pois se trata de um único caso a ser estudado – que é a gestão de dados de pesquisa científica na UFBA –, tendo, como delimitação de análise do “caso” (unidade múltipla de análise), o professor do quadro permanente dos programas de pós-graduação. Único porque exige investigação cuidadosa do caso potencial, com o intuito de minimizar os equívocos apresentados e maximizar o acesso necessário à coleta da evidência do estudo de caso. Seguiram-se todas as etapas apresentadas. Segundo Yin (2015), o método é aplicado em estudos com abordagem tanto qualitativa quanto quantitativa, bem como dos tipos

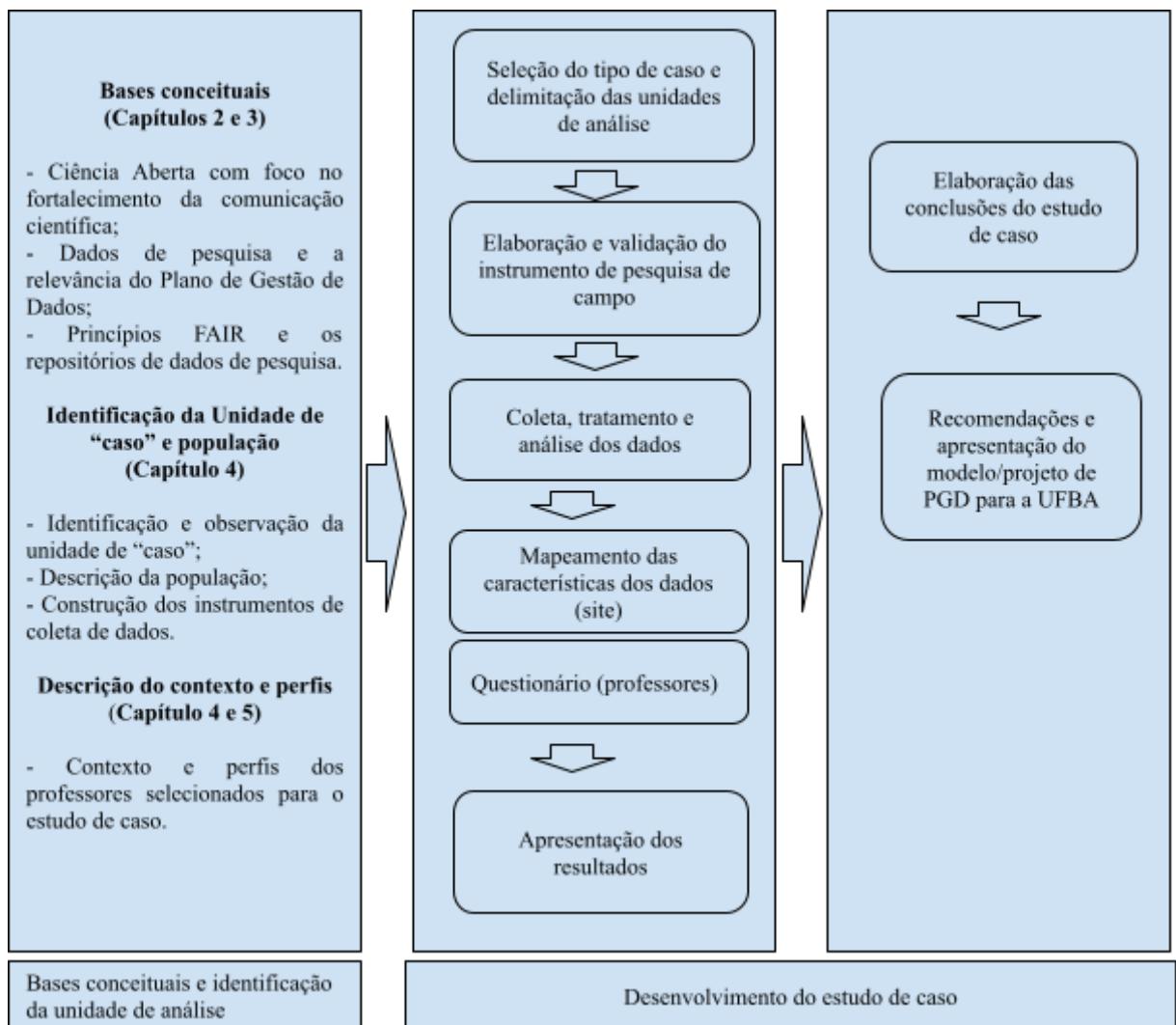
exploratória ou descritiva, pois possibilita a triangulação entre métodos e a utilização de diversas técnicas para a coleta de dados.

Considerando a questão que norteia este estudo – como a Universidade Federal da Bahia tem desenvolvido as ações, no contexto da Ciência Aberta, para a implantação de políticas de gestão de dados de pesquisa, e quais as percepções e os desafios dos professores nesse processo? – e as etapas de realização de um estudo de caso supracitadas, pretende-se responder algumas questões específicas:

- a) Quais iniciativas e práticas de acesso aberto à informação científica a UFBA tem desenvolvido?
- b) Quais são as características dos dados de pesquisa desenvolvidos na UFBA?
- c) Quais obstáculos têm impactado e, consequentemente, distanciado o pesquisador do compartilhamento e abertura de dados de pesquisa?
- d) Qual a percepção dos pesquisadores para a abertura e compartilhamento de dados de pesquisa?

A fim de responder tais questões, elaborou-se um protocolo, baseado na concepção de Yin (2005), para a realização do estudo de caso único integrado.

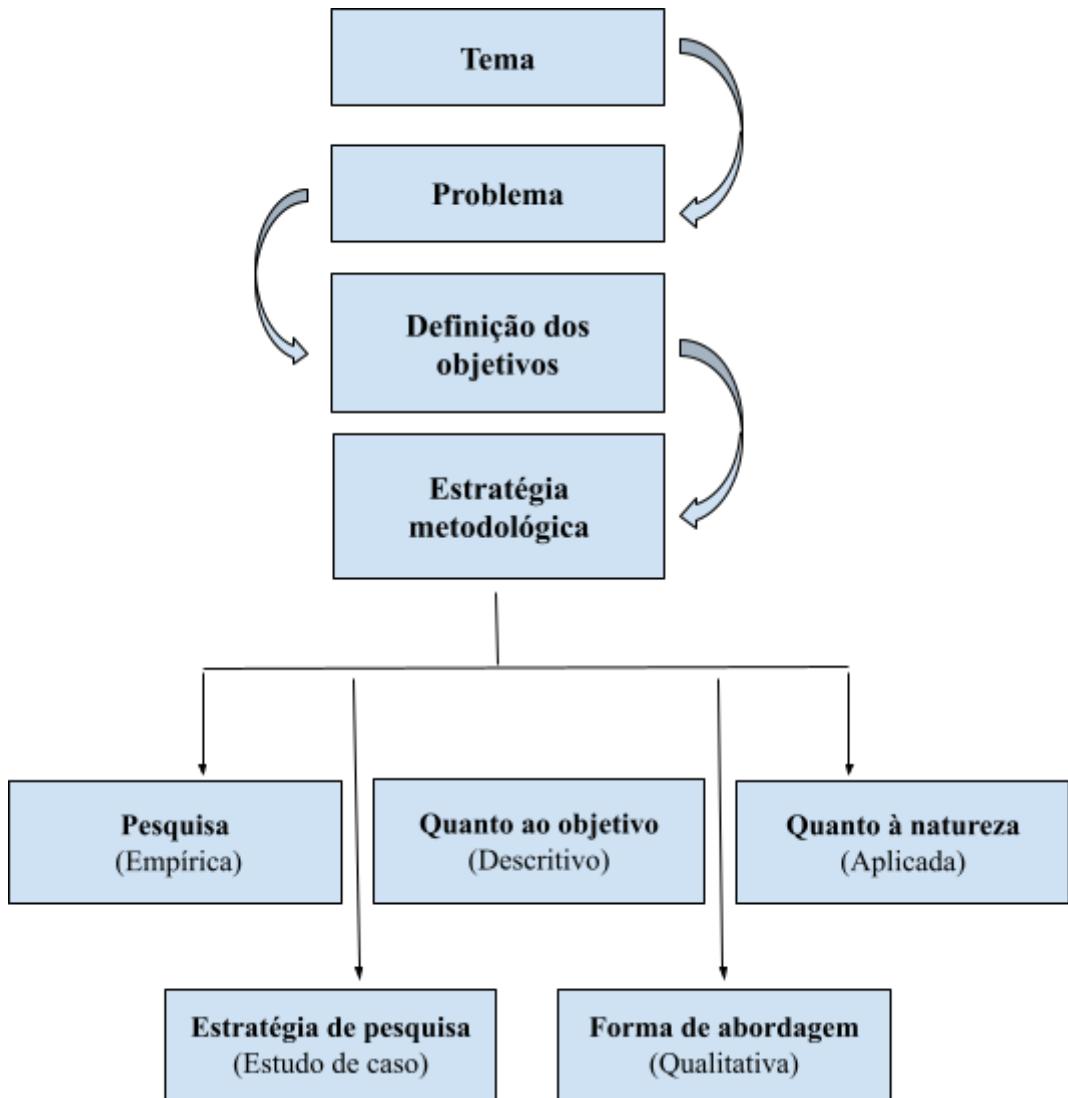
Figura 8 - Protocolo para a coleta de dados



Fonte: Elaboração própria.

A partir dos resultados preliminares da literatura referente a dados de pesquisa e todos os termos correlatos encontrados nas fontes apontadas acima, foi possível compreender a base epistêmica a respeito do objeto de estudo, definir o tipo de pesquisa, bem como os instrumentos para a coleta de dados e apresentar uma proposta de PGD à UFBA. Assim, a Figura 9 resume o delineamento da pesquisa, subsidiando melhor compreensão nas escolhas metodológicas adotadas neste estudo.

Figura 9 - Delineamento da pesquisa e estratégia metodológica



Fonte: Elaboração própria.

Mediante o estudo de caso, será possível verificar ações de melhoria, se necessário, na gestão de dados de pesquisa. Para tanto, as possíveis ações identificadas serão comparadas às etapas de coleta, armazenamento, recuperação e descarte do ciclo de vida dos dados aqui apresentados e com os princípios FAIR.

4.3.2 Elaboração do instrumento de coleta de dados

Com a finalidade de responder aos objetivos desta pesquisa, os procedimentos ocorreram nas etapas apresentadas no Quadro 7. Nas primeira e segunda etapas, os dados foram coletados no *site* da UFBA e por pesquisa bibliográfica; da terceira etapa em diante foi por meio do instrumento de coleta de dados, isto é, o questionário e análise do Plano de

Dados Abertos da Universidade Federal da Bahia.

Quadro 7 - Etapas para a coleta de dados

Objetivos específicos	Etapas	Procedimento de coleta de dados	População-alvo
Identificar e descrever as políticas que podem colaborar com a gestão de dados de pesquisa científica na UFBA	1 ^a	Levantamento do número de programas no site da UFBA	Programas de Pós-Graduação
	2 ^a	Seleção dos programas com conceito igual ou maior que 4	
Conhecer a percepção dos professores pesquisadores da UFBA	3 ^a	Levantamento do quantitativo de professores permanentes em cada programa que atenderam ao critério na etapa 2	Professores permanentes
	4 ^a	Construção do questionário	
	5 ^a	Teste piloto	
	6 ^a	Aplicação do questionário	
	7 ^a	Análise do questionário	
Identificar os principais desafios e as soluções inerentes a construção de uma política estruturada para a gestão de dados de pesquisa científica	8 ^a	Análise do questionário e site da UFBA	Professores permanentes e site da UFBA
	9 ^a	Análise do PDA da UFBA	

Fonte: Elaboração própria.

A análise sistemática da literatura permitiu a elaboração de um conjunto de categorias preliminares acerca do objeto de estudo. A ordenação dessas categorias foi realizada com base nas ordenações conceituais abordadas no referencial teórico. Essas categorias subsidiaram não apenas o questionário, mas toda análise feita no *site* da Universidade.

A construção do questionário foi composta com as seguintes seções: (a) apresentação dos pesquisadores, tema e objetivo da pesquisa, relevância da participação do respondente, incluindo o Termo de Consentimento Livre e Esclarecido (Apêndice A) e os procedimentos para responder ao questionário; (b) informações gerais do respondente e acadêmica (idade, área de conhecimento e tempo de trabalho na instituição); e, (c) as proposições a serem respondidas, compostas por questões fechadas com foco no objetivo do estudo.

Nesse sentido, definiu-se como ferramenta de apoio para a construção e aplicação do questionário, o *Google Forms* – ferramenta utilizada para a criação de formulário on-line. Após a elaboração, o *link* do instrumento foi enviado para os respectivos e-mails dos professores do quadro permanente constantes no *site* da UFBA e no currículo *lattes*, coletados

previamente. A coleta ocorreu entre os meses de outubro e dezembro de 2022. Optou-se por essa modalidade por questão de conveniência, rapidez na entrega da resposta e segurança epidemiológica em virtude da pandemia da Covid-19.

Esse instrumento seguiu o modelo da Escala Likert, escala unificada que possibilita identificar o sentido e a intensidade da atitude. É uma escala de classificação somatória, com o objetivo de expressar atitudes favoráveis ou não, em relação ao objeto de interesse, cujo formato é muito aceito entre os pesquisadores e profissionais do mercado (SANCHES *et al.*, 2011). Esse tipo de escala possibilita combinar um conjunto de respostas dos questionados em relação ao objeto em estudo, a fim de determinar qual a sua posição na escala de atitude desenvolvida. Dessa forma, é possível compreender uma série de afirmações relacionadas ao objeto pesquisado.

Criada pelo norte-americano Rensis Likert, em 1932, a escala – que recebe o nome do criador – é uma das ferramentas mais utilizadas com intuito de medir as atitudes por meio de classificação somatória.

Mediante esse tipo de escala, é possível medir atitudes e conhecer o grau de conformidade do entrevistado com qualquer assertiva proposta. Assim, as respostas servem para obter a intensidade dos “sentimentos” dos questionados. A escala Likert, em geral, possui cinco, sete ou nove níveis, sendo o cinco e o sete os mais utilizados. Para cada nível haverá um número, e será atribuído um escore a cada resposta, sendo 1 a proposição menos favorável e 5 a mais favorável em relação à proposição realizada.

A partir das ideias de Likert, a validação da escala foi um dos aspectos que mais impulsionou. O seu surgimento ocorreu devido à necessidade de responder à principal questão da área, que era como saber se a escala elaborada tinha a capacidade de mensurar o construto desejado. Nessa concepção, “[...] validação com uso de técnicas estatísticas é o nome dado ao conjunto de procedimentos utilizados para conferir maior credibilidade ao processo de mensuração.” (LUCIAN; DORNELAS, 2015, p. 160). Com isso, a academia começou a focar no desenvolvimento de protocolos de elaboração de escalas, descritivos e explicativos, a fim de dar robustez a métodos e técnicas de pesquisa. Para Lucian e Dornelas, (2015, p. 160),

[...] a elaboração de escalas de mensuração envolve a construção de um instrumento em si e a associação de conceitos qualitativos com as métricas quantitativas, ou seja, a atribuição de números a objetos segundo alguma regra determinada (Pooja & Sagar, 2012), a qual busca disciplinar o estudo do fenômeno.

As proposições foram dimensionadas em dez categorias, a saber: a) Ciência Aberta; b) gestão de dados de pesquisas científicas; c) atitude e valores em relação à abertura dos dados; d) motivação para a abertura de dados; e) necessidades dos pesquisadores; f) agência de fomento; g) direito autoral/proteção de dados; h) profissional capacitado e/ou biblioteca; i) razões para não compartilhar dados de pesquisa; e j) repositório de dados.

Na construção desse tipo de escala, obtém-se opções de múltiplas escolhas que incluem, por exemplo, “concordo totalmente”, “concordo parcialmente”, “indiferente”, “discordo parcialmente” e “discordo totalmente”, denominado de diferencial semântico (SANCHES; MEIRELES; DE SORDI, 2011). Cada participante escolhe entre os cinco níveis de concordância ou discordância da proposição, atribuindo, assim, um escore para cada resposta com possível valor, sendo 1 a proposição menos favorável e 5 a mais favorável em relação à pergunta realizada.

Posteriormente, foi realizada a análise dos dados. Para obter uma análise fidedigna, se fez necessário relacionar às proposições estabelecidas com a categoria pretendida e o referencial teórico, conforme ilustra o Quadro 8. Este se apresenta da seguinte forma: quatro colunas, sendo que a primeira se refere às variáveis/categorias; a segunda trata-se da finalidade da categoria; a terceira coluna é dedicada às proposições; e a última relaciona-se com o referencial teórico.

Quadro 8 - Relação das proposições conforme a categoria pretendida e sua relação com o referencial teórico

Variáveis/Categorias	Finalidade	Proposição	Referencial teórico
Ciência Aberta	Conhecer a opinião a respeito de gestão de dados científicos e Ciência Aberta.	1 - Compreendo a relevância da Ciência Aberta para o avanço da ciência. 2 - Devido ao grande volume de dados produzidos na minha pesquisa, utilizo ferramentas computacionais denominadas sistemas de gerência de <i>workflows</i> científicos (por exemplo, o Kepler e o Taverna) a fim de facilitar ou automatizar a gestão dos meus dados e, principalmente, sustentar todo o ciclo de pesquisa.	ALBAGLI; CLINIO; RAYCHTOCK, 2014; GRAY, 2009; SALES, 2014
Gestão de Dados de pesquisa	Conhecer opinião a respeito de gestão de dados de pesquisa científica.	3 - Preciso de fonte de dados confiáveis para o projeto de pesquisa. 4 - A minha instituição possui infraestrutura para a gestão dos meus dados. 5 - Faltam diretrizes para a reutilização de dados na minha instituição. 6 - Percebo dificuldades para implementação da política de gestão de dados científicos na minha Instituição.	SILVA, 2019; CUNHA; COSTA, 2020 GOLD, 2007 COSTA, 2007
Atitude e valores em relação a abertura dos dados	Conhecer atitude e valores a respeito da abertura de dados de pesquisa científica, ou seja, compartilhar, armazenar e preservar os dados em um repositório de dados.	7 - Acredito que aumenta a credibilidade da minha pesquisa. 8 - Contribui para o fortalecimento do avanço científico já que esses dados podem ser utilizados por outros pesquisadores em suas pesquisas. 9 - Contribui com a transparência na conduta da pesquisa. 10 - A abertura de dados de pesquisa contribui para uma maior integridade na pesquisa. 11 - Contribui para uma reproduzibilidade da pesquisa.	GRIGG, 2015
Motivação para a abertura	Conhecer os motivos para abertura ou não dos dados científicos coletados pelo	12 - Amplia a acessibilidade da minha pesquisa, garantindo que o material seja acessível de maneira estável por longo tempo.	GRAY, 2007 HEY; TREFETHEN, 2003

	pesquisador durante a fase da pesquisa.	13 - Possibilita maior confiabilidade já que o material em repositório tem credibilidade. 14 - Aumenta o prestígio em minha área e, consequentemente, reconhecimento profissional. 15 - Amplia a visibilidade da minha pesquisa.	
Necessidades dos pesquisadores	Agora queremos conhecer as necessidades do pesquisador sobre gestão de dados de pesquisa científica na sua Instituição.	16 - Sou obrigado(a) a abrir por causa da minha agência financiadora. 17 - Apoio técnico na elaboração do plano de gestão de dados, ou seja, identificação de normas e esquema disciplinares de descrição de dados. 18 - Necessito de serviço de armazenamento e preservação, do tipo <i>software</i> que possa contribuir no armazenamento e preservação dos meus dados. 19 - Necessito de <i>backup</i> de arquivos. 20 - Preciso de formação e/ou consultoria na gestão de dados de pesquisa.	FIGUEIREDO, 1979, 1998 PULJAK <i>et al.</i> , 2022
Agências de Fomento	A próxima sequência de perguntas refere-se à opinião do pesquisador a respeito da exigência por parte de algumas agências de fomento quanto à apresentação do plano de gestão de dados de pesquisa.	21 - A agência de fomento exige um plano de gestão de dados no projeto de sua pesquisa, mas não sei elaborar esse documento. 22 - Compartilho os dados para atender a demanda da agência de fomento.	MEDEIROS, 2016
Direito autoral/ Proteção de dados	Para reutilização e compartilhamento dos dados é necessário que o pesquisador conheça sobre direito autoral e proteção de dados.	23 - Meus dados são sensíveis (potencialmente sigilosos). 24 - Meus dados são ostensivos (sem sigilo). 25 - Meus dados podem ou geram patentes. 26 - Meus dados são pessoais.	BRASIL, 1998 LIMA; SANTINI, 2008

		27 - Meus dados têm parceria interinstitucionais.	
Profissional capacitado e/ou Biblioteca	Para o gerenciamento dos dados de pesquisa, a presença de um consultor deve contribuir para uma melhor agilidade ou não no processo de gestão. Sobre isso, a série de questões abaixo está relacionada à necessidade de um profissional capacitado para o assessoramento, junto ao pesquisador, no processo de gestão dos dados de pesquisa.	28 - Sinto necessidade de profissional capacitado para orientar em todo processo da gestão dos meus dados científicos, desde a gênese até o armazenamento e preservação. 29 - Percebo que a minha Instituição tem contribuído para a formação de profissional capaz de realizar o gerenciamento de dados científicos. 30 - A Biblioteca poderia gerenciar os meus dados científicos por meio de um repositório de dados. 31 - Não é papel da Biblioteca gerenciar os meus dados e sim eu, pesquisador. Cabe uma Assessoria fornecida pela Pró-reitoria.	FIGUEIREDO, 1979, 1998
Razões para não compartilhar dados de pesquisa	As perguntas a seguir estão voltadas aos motivos para efetivação do compartilhamento dos dados de pesquisa.	32 - Não sei como compartilhar dados. Preciso de suporte (assessoria). 33 - Meus dados não podem estar abertos, pois contêm informação sigilosa. 34 - Mau uso dos dados pode causar uma má interpretação por outros pesquisadores. 35 - Se eu compartilhar, meus dados podem ser usurpados por outros pesquisadores.	PULJAK <i>et al.</i> , 2022
Repositório de dados	Conhecer a opinião sobre repositório de dados de pesquisa.	36 - Não confio na transparência e na autenticidade dos dados brutos em repositório de dados. 37 - Não tenho motivos/vantagens que incentivam a publicação de dados em um repositório de acesso aberto. 38 - Necessito de serviços de compartilhamento e acesso a dados. 39 - Não tenho conhecimento de nenhum repositório para fazer o compartilhamento.	MONTEIRO, 2021 PAVÃO <i>et al.</i> , 2018

		40 - Não compartilho porque meus dados têm restrição contratual. 41 - Não compartilho porque posso perder uma boa publicação, caso eu compartilhe os dados	
--	--	--	--

Fonte: Elaboração própria.

Após o cumprimento e aplicação do instrumento, a análise foi realizada e apresentada no Capítulo 5. O instrumento de coleta de dados utilizado procurou pormenorizar a problemática da gestão de dados, compreender o lugar da UFBA na construção de políticas para tal gestão no estado, e contribuir no desenvolvimento de um esboço de diretrizes para a gestão desses dados a partir de diferentes perspectivas – funcionários de agências de fomentos *versus* pesquisadores, pesquisadores *versus* acesso aberto, pesquisadores *versus* sociedade.

4.3.2.1 Questionário - teste piloto

Após a elaboração dos instrumentos de pesquisa, foi necessária a realização de um método de validação do instrumento de coleta de dados. Esse processo de validação – que no caso foi o questionário em escala Likert – consiste em estabelecer critérios a fim de verificar a qualidade do instrumento, se mensura o que realmente se propõe, e refere-se a uma análise detalhada do instrumento, com o objetivo de verificar o item proposto (ALEXANDRE; COLUCI, 2011; CUNHA; ALMEIDA NETO; STACKFLETHC, 2016).

O teste piloto foi realizado com a finalidade de verificar o processo metodológico utilizado na pesquisa e ratificar os instrumentos para a execução do estudo. Para aplicação, foi selecionada uma amostra de professores permanentes dos cursos de pós-graduação coletada no *site* da Universidade Estadual da Bahia (UNEBC) composta por 100 professores de cinco cursos. A seleção dos cursos foi obtida aleatoriamente, contudo os professores pertencentes aos cursos teriam que ter os seus respectivos e-mails constantes no *site* já que nem todos os cursos constavam a relação dos e-mails dos professores permanentes.

A Universidade do Estado da Bahia foi escolhida por ser a maior instituição pública de ensino superior da Bahia. A Instituição foi fundada em 1983 e é mantida pelo Governo do Estado por intermédio da Secretaria da Educação (SEC) e está presente geograficamente em todas as regiões do Estado, estruturada no sistema multicampi. Atualmente, a universidade disponibiliza mais de 170 cursos e habilitações nas modalidades presencial e de educação a distância (EaD), nos níveis de graduação e pós-graduação, oferecidos nos 30 departamentos. Cabe frisar, o expressivo crescimento na oferta de cursos stricto sensu (mestrados e doutorados) nos últimos anos, em Salvador e outras cidades, promovendo a interiorização da pós-graduação pública, gratuita e de qualidade.⁴⁷

⁴⁷ Ver mais informações em www.uneb.br

A UNEB possui programas de pós-graduação em todas as áreas do conhecimento. Esses cursos podem ser *stricto sensu*, como são conhecidos os mestrados e doutorados, ou *lato sensu*, que é o caso das especializações, a saber: mestrado acadêmico (13 cursos), mestrado profissional (9 cursos), doutorado (4 cursos), e especialização (7 cursos).

O questionário foi composto por 44 perguntas; as três primeiras correspondem ao perfil do respondente e as demais em número de 41 se referem à temática da pesquisa. O teste piloto foi aplicado com enfoque crítico para identificar deficiências e imprecisões nos instrumentos utilizados. Além disso, buscou- se verificar a pertinência dos resultados obtidos na pesquisa com os objetivos do trabalho e aplicar o teste mais coerente para o plano de análise de dados para os resultados do público-alvo que seria os professores permanentes da UFBA. Sendo assim, apenas os principais resultados serão apresentados nessa seção.

Em outubro de 2022, 100 e-mails foram enviados com prazo de 15 dias para que os professores respondessem. Não obtendo o quantitativo desejável e finalizado o prazo inicial foi encaminhado para cada coordenação dos cursos selecionados uma solicitação de contribuição na realização da pesquisa, nesse caso, com o prazo de 10 dias. Dos 100 e-mails enviados, apenas 10% responderam, quantitativo bem abaixo do esperado, mas para a finalidade do teste foi adequado para avaliar as posições falhas e incoerências. Durante a realização do teste-piloto, algumas adequações foram realizadas para aperfeiçoar o questionário, tais como, pequenos ajustes na formatação do instrumento e reformulação das proposições 2 e 16. Na segunda proposição foi acrescentado exemplos de workflows científicos; e na décima sexta, foi correção ortográfica.

Considerando o público-alvo (os professores de pós-graduação do quadro permanente) e o instrumento de coleta de dados (o questionário) com o uso da escala Likert de 5 pontos – contendo assertivas e colunas que permitirão indicar o grau de concordância da proposição –, os dados foram analisados e validados por meio da Lógica Paraconsistente (LP).

A LP é um método indicado para pesquisas qualitativas em Ciências Sociais Aplicada, capaz de unificar informações oriundas da literatura científica (objetivas) e provenientes de valores e opiniões de especialistas em diversas áreas do conhecimento (subjetivas). Além disso, essa técnica permite lidar com dados incertos e informações contraditórias sem se tornar trivial (DA COSTA *et al.*, 1999), ideia que também pode ser estendida para conjunto de informações, conjunto de dados, entre outros. Com base em Sanches, Meireles e De Sordi (2011), o uso da Lógica Paraconsistente possibilita analisar dados complexos de forma mais

simplificada, relacionado às implicações existentes entre a conclusão e a premissa que serve de apoio.

Analizar dados oriundos de opiniões e atitudes não é uma tarefa fácil, pois são variáveis que dependem da ação e reação das pessoas. O uso da escala Likert pode diminuir essa dificuldade pois destina-se a quantificar essas opiniões e atitudes, contudo exige cautela pois essa escala contém dados ordinais que admitem apenas teste não paramétricos com medidas estatísticas, a saber: centralidade, dispersão, associação ou correlação, significância e aderência. Diante disso, e em decorrência da abordagem qualitativa deste estudo, o uso da Lógica Paraconsistente se faz necessário para se obter uma interpretação e síntese de informação por meio da escala tipo Likert.

Cabe lembrar que os dados oriundos desse tipo de escala possibilitam apenas a relação de equivalência (=) e de comparação (>) o que dificulta a análise dos dados. A LP, portanto, interpreta e sintetiza a informação obtida por essa escala. Sendo assim, essa técnica se aplica por conferir ao trabalho uma “robustez demonstrativa”; ou seja, modela conhecimentos mediante a procura por evidências, de tal maneira que os resultados obtidos se aproximam do raciocínio humano (SANCHES; MEIRELES; DE SORDI, 2011).

As escalas utilizadas em pesquisas qualitativas com a finalidade de medir opiniões atitudes possuem pontos que são denominados de diferencial semântico. Sendo assim, todo o cálculo foi realizado com base nas orientações do método paraconsistente de Sanches, Meireles, De Sordi (2011), e nos valores apresentados no diferencial temático (DT, DP, I, CP e CT) da escala Likert. O conceito de diferencial semântico sinaliza que as expressões utilizadas variam de sentido a depender da natureza da pesquisa. Logo, os respondentes devem escolher entre diversas opções e marcar aquela que mais se aproxima da sua opinião ou atitude.

A escala Likert apresentada neste estudo possui cinco pontos, o ideal para uma análise mais robusta. O cálculo se dá da seguinte forma: primeiro, somam-se os valores marcados de cada respondente em cada item de cada categoria e, posteriormente, divide-se pelo total de números de itens. Depois, somam-se as médias de cada respondente em cada fator e divide-se pelo número de respondentes e, assim, sucessivamente (SANCHES; MEIRELES; DE SORDI, 2011).

Para saber o número de concordantes das proposições (Cp), é dado o seguinte algoritmo: $Cp = \sum CP + \sum CT + \sum I * 1/2$. De forma semelhante, é obtido o cálculo de discordantes

das proposições ($D_p = \sum DP + \sum DT + \sum I * 1/2$). Em seguida, o indicador de força relativa indica o Grau de Concordância da proposição (GC_p), que é obtido pelo seguinte cálculo: GC_p = 100 - (100/ (Cp/Dp+1)). Esse cálculo é feito com o intuito de dirimir erros de divisão. Com a finalidade de interpretar os valores da coluna do GC_p, Sanches, Meireles, De Sordi, (2011) sugerem os parâmetros apresentados por Davis (1976), conforme o Quadro 9. A crença e descrença de que o conjunto das proposições são verdadeiras é dado por $\mu_2 = \sum QT/Cp$ crença e $\mu_1 = \sum QT/Dp$ (SANCHES *et al.* 2011), onde QT é a quantidade total de respondentes.

Quadro 9 - Interpretação de valores de GC_p

Valor de GC	Frase Adequada
90 ou mais	Uma concordância muito forte
80 a + 89,99	Uma concordância substancial
70 a + 79,99	Uma concordância moderada
60 a + 69,99	Uma concordância baixa
50 a + 59,99	Uma concordância desprezível
40 a + 49,99	Uma discordância desprezível
30 a + 39,99	Uma discordância baixa
20 a + 29,99	Uma discordância moderada
10 a + 19,99	Uma discordância substancial
9,99 ou menos	Uma discordância muito forte

Fonte: Sanches, Meireles e De Sordi (2011).

A análise dos dados com base na LP possibilita estabelecer graus de crença e descrença de uma proposição, pois “[...] pode ser aplicada para modelar conhecimentos de procura de evidências de tal forma que os resultados obtidos são aproximados do raciocínio humano [...]” (DA COSTA *et al.*, 1999, p. 37). Da Costa e colaboradores (1999, p. 19) afirmam que, na análise da LP, “[...] as anotações são representativas de graus de crença (μ_1) e descrença (μ_2) atribuídos à proposição, dando-lhes conotações de valoração [...].”

A LP funciona da seguinte forma: a) se existir um alto grau de contradição, não existe consistência quanto ao achado, portanto, devem-se buscar novas evidências; b) se existir um baixo grau de contradição, pode-se formular a conclusão, desde que se tenha um alto grau de certeza. Assim, Da Costa (1999, p. 37) mostra que a LP, além de lidar com a certeza, lida também com contradições e indeterminações, a fim de “[...] demonstrar que a Lógica Paraconsistente pode ser aplicada para modelar conhecimentos por meio de procura de evidências de tal forma que os resultados obtidos são aproximados do raciocínio humano”.

Efetuada essa parte a etapa seguinte foi analisar a associação entre as proposições de cada categoria, para tanto foi aplicado o coeficiente de correlação de Spearman (FÁVERO;

BELFIORE, 2017) que mensurou a associação entre as proposições de cada categoria (Quadro 8). Ressalta que o nível de significância adotado foi de 5%.

O coeficiente de correlação de Spearman (r_s) avalia com que intensidade a relação entre duas proposições pode ser descrita pelo uso de uma função monótona, ou seja, quando as variáveis tendem a mudar juntas, mas não necessariamente a uma taxa constante, ou melhor, uma função monótona é aquela que nunca diminui ou nunca aumenta, já que é um aumento variável independente

A correlação de Spearman é calculada a partir da seguinte fórmula:

$$r_s = 1 - \frac{6 \sum d_i^2}{n(n^2-1)}$$

Onde:

n = número de pontos de dados das duas variáveis;

d_i = diferença de alcance do elemento “n”

O coeficiente de Spearman (r_s), pode assumir um valor entre -1 e +1; um valor de +1 em significa uma associação de classificação perfeita. Valor de 0 significa que não há associação de classificação, já um valor de -1 significa uma associação negativa perfeita. Se o valor se aproxima de 0, a associação entre os dois intervalos é mais fraca ao passo que quando se aproxima de 1 a associação é mais forte. Para o cálculo do r_s foi utilizado o Statistical Package for the Social Science (SPSS), um pacote estatístico, desenvolvido pela International Business Machines Corporation (IBM), versão 22 (FIELD, 2009).

Como foi dito anteriormente o r_s varia de -1 a +1 e o resultado gera vários valores. Para sua interpretação foi usado a proposta por Rumesy (2019, p. 313), descrita abaixo:

- Exatamente **-1**. Uma relação linear descendente (negativa) perfeita
- **-0,70**. Uma forte relação linear descendente (negativa)
- **-0,50**. Uma relação descendente moderada (negativa)
- **-0,30**. Uma relação linear descendente (negativa) fraca
- **0** Nenhuma relação linear
- **+0,30**. Uma relação linear ascendente fraca (positiva)
- **+0,50**. Uma relação ascendente moderada (positiva)
- **+0,70**. Uma forte relação linear ascendente (positiva)
- Exatamente **+1**. Uma relação linear ascendente (positiva) perfeita

4.3.3 Procedimentos de coleta de dados

Para cada objetivo específico, os procedimentos de coleta de dados selecionados para compor a pesquisa, necessitou utilizar estratégia adequada, conforme descrito no Quadro 6.

A coleta de dados teve início após a elaboração e validação do instrumento, entre outubro e dezembro de 2022. O questionário foi aplicado eletronicamente por meio da plataforma *Google Forms* para os respectivos e-mails constantes no *site* da UFBA e no currículo *lattes*. Optou-se por essa modalidade por questão de conveniência, rapidez na entrega da resposta e segurança epidemiológica em virtude da pandemia da Covid-19.

Os participantes foram convidados a assinar o Termo de Consentimento Livre e Esclarecido (TCLE), no qual constaram os objetivos da pesquisa, confirmando sua aceitação em participar. Em seguida, cada um preencheu o formulário e encaminhou para a pesquisadora.

Após o cumprimento das etapas, os dados foram analisados conforme o item 5.

5 ANÁLISE E DISCUSSÃO DOS DADOS

Neste capítulo são apresentados os resultados da pesquisa que teve o propósito de conhecer a percepção dos professores pesquisadores no que tange à gestão de dados e às ações da UFBA desenvolvidas para a implantação de um plano de gestão de dados de pesquisas no contexto da Ciência Aberta, a qual corresponde aos objetivos específicos: OE2⁴⁸, OE3⁴⁹ e OE4⁵⁰.

Ressalte-se que o OE1 que é verificar os principais conceitos da Ciência aberta, dados de pesquisa, Plano de gestão de dados e FAIR – foi alcançado e apresentado nos capítulos 2 e 3.

5.1 ANÁLISE DAS POLÍTICAS PARA GESTÃO DE DADOS DE PESQUISA NA UFBA

Com o intuito de responder o objetivo OE2 foi feito um mapeamento no *site* da Universidade Federal da Bahia a fim de encontrar ações e/ou legislação ou qualquer ato normativo que abordem sobre gestão de dados e/ou acesso aberto.

Iniciou-se pela legislação, utilizando as seguintes palavras-chaves: “ciência aberta”, “dados de pesquisa”, “dados científicos”, “compartilhamento de dados de pesquisa” e “abertura de dados de pesquisa” nos regimentos, regulamentos, portarias, resoluções e instruções normativas; e constatou-se que não continha nada a respeito do movimento da Ciência Aberta. Os mesmos descritores foram utilizados na busca do site da Universidade. Contudo, os resultados não foram precisos. Mesmo usando as aspas, os resultados não tinham a ver com o tema, mas sim com as matérias institucionais. Destas, apenas três, de fato, estavam relacionadas com a temática em estudo: “Escola Politécnica promove workshop sobre inovação aberta na ACTA-11”, publicada em 31 de outubro de 2011; “Comunidade científica debate percepção pública da ciência”, em 4 de novembro de 2022; e “Plano de Dados Abertos da UFBA está disponível para consulta pública” publicada em 10 de dezembro de 2012.

Posteriormente foi analisado “O Plano de Desenvolvimento Institucional (PDI) da UFBA 2018-2022”⁵¹ com a finalidade de buscar qualquer menção ao movimento. O PDI é o

⁴⁸ OE2: mapear as ações/políticas desenvolvidas para a implantação de uma política de gestão de dados de pesquisa/Plano de gestão de dados;

⁴⁹ OE3: identificar, por meio do questionário, a opinião dos professores pesquisadores fatores que podem influenciar a abertura e/ou compartilhamento de dados de pesquisa científica;

⁵⁰ OE4: descrever os principais desafios e as soluções inerentes à construção de uma política estruturada para a gestão de dados de pesquisa científica.

⁵¹ Disponível em: <https://proplan.ufba.br/sites/proplan.ufba.br/files/pdi-2018-2022.pdf>. Acesso em: 9 mar. 2023.

documento que orienta as políticas e práticas de gestão e autogestão da Universidade. Tem, portanto, caráter estratégico ao projetar a UFBA que se pretende construir, sem desconsiderar sua trajetória histórica e os desafios resultantes do contexto em que atua. Trata-se de um documento que busca conciliar os elementos que definem um PDI (Decreto MEC nº 5.773, de 9 de maio de 2006) com os procedimentos técnicos que embasam o planejamento estratégico de qualquer instituição. Dentre as políticas e práticas de gestão e autogestão, o PGD não é contemplado, inclusive, ao relatar os cenários futuros e as perspectivas para a Universidade pública brasileira a partir dos potenciais e prováveis cenários para o país, no PDI não consta a construção do PGD, mas sinaliza aproximações com o movimento da Ciência Aberta.

No entanto, é interessante abordar o detalhamento das Diretrizes Estratégicas definido no PDI, especificamente o “subeixo Pesquisa”. As diretrizes apresentadas trata-se de ações articuladas para atingir as metas de cada objetivo estratégico. Para isso, estas são divididas em dois grandes eixos que organizam o planejamento da Universidade: o eixo da missão (relativo às atividades fins) e o eixo dos recursos (relativo às atividades meio). É de interesse dessa pesquisa, o eixo da missão onde diz respeito à estrutura do projeto pedagógico institucional, ou seja, a compreensão que se têm de Ensino, Pesquisa e Extensão.

Para esse estudo, o subeixo Pesquisa traz informações relevantes:

- a) papel social: a transferência do que é gerado na Universidade para a sociedade por meio das publicações. E como centros de produção de conhecimento e de geração de tecnologia, as universidades passaram a receber demandas crescentes de apoio à solução de problemas práticos da sociedade. Nesse sentido, deve-se lembrar duas situações que vão ao encontro desse subeixo: a Lei de Acesso à Informação que se deu em decorrência de uma demanda social com a finalidade de regulamentar as informações que estavam sob sigilo; e, a Ciência Cidadã – um dos pilares da Ciência Aberta, porém é preciso que a sociedade participe no processo de investigação.
- b) o fortalecimento da Pesquisa: nesse ponto, busca-se apoio em valores que devem continuar orientando políticas, programas e projetos específicos voltados para fortalecer a atividade científica na UFBA, dentre eles: autonomia do pesquisador na escolha dos seus problemas de pesquisa; fortalecimento de redes de parcerias, fundamentais para o intercâmbio de saberes, o que implica a necessidade de eventos e encontros periódicos da comunidade interessada em

algum tema, entre outros. Nesse caso, os repositórios de dados são uma ferramenta essencial para a troca de saberes.

Como se vê, as informações referentes ao subeixo Pesquisa são de caráter generalista, mas que coadunam com a proposta da Ciência Aberta, porém é necessário abordar com maior afinco as facetas e as questões relacionadas ao acesso aberto e compartilhamento.

É no objetivo estratégico de nº 6 do PDI que a UFBA prevê a implantação de plano de gestão de dados. Embora o PGD não esteja contemplado em detalhes no PDI, o objetivo 6 menciona a implantação de um plano de gestão de dados, o que demonstra um movimento na direção correta, apesar de a abordagem ainda ser genérica e não detalhada.

OBJETIVO ESTRATÉGICO 6

Aprimorar e fortalecer os processos e instrumentos de governança institucional, com o fomento à integração, articulação e sinergia das atividades de ensino, pesquisa e extensão e o *estímulo à participação cidadã*, com transparência pública e diálogo com a comunidade universitária e a sociedade (UFBA, 2017, p. 103, grifo nosso).

Na especificação desse objetivo, constam as diretrizes estratégicas, bem como as descrições sintéticas das ações e as metas específicas para 2022, conforme consta no Quadro 10.

Quadro 10 - Diretrizes estratégicas do PDI da UFBA, 2018-2022

DIRETRIZES ESTRATÉGICAS	DESCRÍÇÃO SINTÉTICA DAS AÇÕES	METAS ESPECÍFICAS - 2022
Prover transparência ativa, sustentabilidade, satisfação e participação do público e comunidade UFBA com a implantação do Plano de Dados Abertos na Universidade e Tramitação Eletrônica de Processos Institucionais.	Atuar em conformidade com a Política de Dados Abertos, instituída pelo Governo federal, e com a Lei de Acesso à Informação (LAI). A externalização dos dados da Universidade valoriza a transparência, participação ativa, controle social, e retroalimentação de insumos para o alcance de benefícios para a sociedade e academia. Diretriz em conformidade com o Processo Eletrônico Nacional (PEN), programa do Governo federal que, através da tramitação eletrônica de processos, visa agilidade, produtividade, satisfação do público usuário, redução de custos, sustentabilidade, compartilhamento de conhecimento, transparência ativa, mudança organizacional e qualidade de vida do servidor.	Plano de dados abertos implantado Processo Eletrônico implantado, com conjunto expressivo de processos institucionais em tramitação eletrônica.

Fonte: UFBA, 2017, p. 103.

A implantação do Plano de Dados Abertos (PDA) na Universidade Federal da Bahia sinaliza uma etapa bastante significativa para o fortalecimento da ciência na instituição. Como meta a ser alcançada, em 2022 a UFBA finalizou o PDA. Para essa etapa final, houve fases em que foi encaminhado para a consulta pública. Conforme matéria publicada no site da UFBA, em 10 de dezembro de 2018, a minuta do Plano de Dados Abertos da Instituição para os anos de 2019-2020 ficou disponível para as observações da comunidade universitária e sociedade em geral, durante 15 dias, a partir do dia 10 de dezembro a 24 de dezembro de 2018.

O plano foi elaborado por uma Comissão de Dados Abertos e, em consonância com o PDI, objetiva com a abertura dos dados da Universidade promover a visibilidade e a transparência de informações, atividades, ações e serviços da UFBA; além disso ampliar a transparência ativa na Administração Pública e na Ciência Aberta, com o compromisso de divulgação permanente de dados públicos em formato aberto “manuseável”, pelo usuário, afirma Fabíola Greve, a assessora especial de tecnologia da informação do gabinete da Reitoria à época.

Eduardo Mota, pró-reitor de planejamento da UFBA, informa na matéria que

os dados abertos disponibilizados pela UFBA serão coletados, prioritariamente, a partir de informações consolidadas nos sistemas de informação da instituição, depois serão periodicamente atualizados, preferencialmente de forma automática. Dessa forma, assume-se o compromisso de divulgação permanente dos dados de interesse público produzidos nas diversas ações realizadas (UFBA, 2018, on-line).

Há de convir que dados aberto é diferente de transparência. Algumas administrações públicas a fim de cumprir a Lei de Acesso à Informação (LAI) e de “aderir” ao movimento da Ciência Aberta, pensa que dar transparência às informações públicas, ou seja, publicizar, é disponibilizar os dados de forma abertamente, o que parece ter sido o equívoco do pró-reitor Mota.

A LAI possibilitou uma cultura de governo aberto, garantindo à sociedade o direito de obter a informação pública e obrigando o Estado a fornecer a transparência passiva, – ato de responder às demandas informacionais da sociedade – e a transparência ativa, ato de publicar a informação sem solicitação prévia e, neste caso, a divulgação de informações públicas é feita em seus sítios na Internet (BRASIL, 2011, art. 3º). A Lei nº 12.527/2011 prevê dados abertos em seu artigo 8º, mas é a Política Nacional de Dados Abertos, por meio do Decreto nº

8.777/2016 que o institui, frisando que dados abertos inclui transparência, mas vai além disso. Talvez seja essa a confusão (SENA, 2019).

Ainda sobre o plano de dados abertos implantado pela UFBA, no *site* não é muito fácil de achar. No *link* “Dada abertos” não direciona para o acesso ao plano, tampouco para o portal de dados abertos que consta no *link*, isso mostra uma falha do sistema. O mesmo só foi encontrado no *site* de busca “Google” (ANEXO A). Isso, talvez, justifique a falta de conhecimento dos pesquisadores sobre esse plano.

Outra ação importante é o repositório institucional⁵² que tem por objetivo reunir num único local virtual a produção acadêmica da Universidade, contribuindo para ampliar a visibilidade da Instituição e dos seus pesquisadores, assim como o impacto da investigação, além da preservação da memória intelectual, seja na área das artes, das ciências, humanidades, da tecnologia, da inovação ou da criação. Mas é bom lembrar que repositório institucional é diferente de repositório de dados, conforme consta na Seção 3.2.3 deste estudo.

Por fim e após exaustiva análise no site, não foi possível identificar políticas que direcionem a gestão de dados de pesquisa, objeto deste estudo.

Como visto na fundamentação teórica deste estudo, os dados produzidos como parte das pesquisas têm uma ampla variedade de formatos, por esse motivo, a literatura apresenta inúmeras propostas de ciclo de vida dos dados, todavia, as etapas sequenciais da gestão dos dados representam as atividades que o pesquisador precisará executar para que os seus dados de pesquisa sejam disponibilizados e precisam ser incorporar à própria metodologia da pesquisa. Sendo assim, utilizou-se a proposta da Figura 4 acatada como viés condutor para construir um projeto piloto de gestão de dados de pesquisa para a UFBA (APÊNDICE C).

5.2 ANÁLISE DO QUESTIONÁRIO

Para os demais objetivos específicos (OE3 e OE4), foi utilizado um questionário que foi enviado para o público-alvo. O instrumento foi composto por 44 proposições: 3 qualitativas nominais e 41 qualitativas ordinais. A primeira parte procurou delinear um perfil do respondente, referindo-se aos dados demográficos e continham as seguintes variáveis: área de conhecimento, faixa etária e tempo de trabalho na instituição. Por meio dessas perguntas foi possível compreender a correlação entre o compartilhamento ou não dos dados com as áreas de pesquisa bem como com o tempo de instituição.

⁵² <https://repositorio.ufba.br/>

Para averiguar a consistência interna do questionário foi aplicado o Coeficiente Alfa de Cronbach. O valor do coeficiente alfa encontrado, considerando-se todas as seções, foi 0,8016, valor considerado de confiabilidade alta, segundo parâmetros utilizados por Freitas e Rodrigues (2005). Cabe lembrar que o cálculo da precisão da escala também foi realizado por meio do coeficiente alfa de Cronbach.

Em seguida passou-se a caracterizar os respondentes. Vale ressaltar que da população-alvo de 880 professores do quadro permanente com os critérios preestabelecidos e descritos nos procedimentos metodológicos, responderam ao questionário 83 (9,43%), sendo que o maior número de respondentes coube aos pesquisadores da área de Ciências Humanas, equivalente a 49,4%, seguido de Ciências Biológicas/Saúde (21,69%) e Exatas (19,28%), na Tabela 2 são apresentados esses demonstrativos.

Tabela 2 - Relação de respondentes por Área de conhecimento

Área de conhecimento	Nº de respondentes	%
Artes	3	3,61
Ciências Biológicas	18	21,69
Ciências Exatas	16	19,28
Ciências Humanas	41	49,4
Letras	5	6,02
Total	83	100

Fonte: Dados da pesquisa.

Além da área de conhecimento fez parte das perguntas a relação do tempo de trabalho na Instituição de Ensino e a faixa etária dos professores. Quanto à faixa etária dos professores respondentes, os resultados mostraram que a maioria dos deles possui mais de 51 (55,42%) anos (Tabela 3).

Tabela 3 - Relação de professores por faixa etária

Idade	Nº de respondentes	%
29 a 39 anos	9	10,84
40 a 50 anos	28	33,73
51 anos ou mais	46	55,42
Total	83	100

Fonte: Dados da pesquisa.

No que diz respeito ao tempo de serviço na UFBA, podemos observar (Tabela 4) que 42,17% estão entre 11 a 25 anos.

Tabela 4 - Relação de professores por tempo de trabalho na Instituição/UFBA

Tempo de trabalho na Instituição	Nº de respondentes	%
Menos de 10 anos	24	28,92
11 a 25 anos	35	42,17
26 a 36 anos	12	14,46
37 anos ou mais	12	14,46
Total	83	100

Fonte: Dados da pesquisa.

As 41 perguntas foram divididas em 10 seções, de acordo com o eixo que elas mensuraram. Para análise desses dados, utilizou-se da lógica paraconsistente (referência) a fim de mensurar as opiniões dos respondentes. De fato, tornar universal o acesso a dados, principalmente oriundos de recurso público, é um grande desafio para a Ciência Aberta. Óbvio que, além da motivação comum que está na base do movimento pelo acesso aberto à informação científica, é necessário discussão e compromisso com os atores envolvidos nesse processo. Como uma ponta desse *iceberg*, se encontram os pesquisadores, ou seja, os criadores desses dados. Para tanto compreender e participar desse movimento é um grande passo com uso de ferramentas que são primordiais no compartilhamento dos dados e no fortalecimento científico.

Baseado na análise dos dados por meio da proposta de Macnaughton e do oscilador estocástico de Wilder Júnior da categoria Ciência Aberta, ao ser questionado sobre a temática, o sentimento comum foi de concordância para a relevância da ciência aberta e dos uso de ferramentas que contribuem para gerir e sustentar o ciclo de pesquisa. As opiniões dos questionados a respeito da Ciência Aberta demonstram que a temática em questão é relevante para o avanço científico, contudo a discordância no uso das ferramentas de gerência enfatiza o pouco ou nenhuma aderência ao movimento de compartilhamento de pesquisas científicas. Esse sentimento corrobora com o valor GCp (63,25), mas indica a existência de uma concordância baixa para as proposições referentes à temática apresentada (Quadro 11).

Conhecer a respeito da CA é fundamental pois ela incentiva o desenvolvimento e uso de plataformas computacionais avançadas de apoio à pesquisa, considerando os direitos de acesso e uso dos dados assim como aqueles resultantes de pesquisa, ou seja, é uma maneira de avançar as fronteiras do conhecimento por meio das tecnologias (GRAY, 2009; SALES, 2014).

Quadro 11 – Ciência Aberta

Proposição Categoria 1: Ciência Aberta	Diferencial Semântico					Total	Discordantes da proposição (D_p)	Concordantes da proposição (C_p)	Grau de Concordância da proposição (GCP)
	1-DT	2-DP	3-I	4-CP	5-CT				
1. Compreendo a relevância da Ciência Aberta para o avanço da ciência	3	2	2	14	62	83	6	77	63,25
2. Devido ao grande volume de dados produzidos na minha pesquisa, utilizo ferramentas computacionais denominadas sistemas de gerência de workflows científicos (por exemplo, o Kepler e o Taverna) a fim de facilitar ou automatizar a gestão dos meus dados e, principalmente, sustentar todo o ciclo de pesquisa.	34	8	26	4	11	83	55	28	
Total	61			105		166	0,3675	0,6325	
	D_p			C_p			μ2	μ1	

Fonte: Elaboração própria.

Entende-se Ciência Aberta como um movimento em processo de construção (ALBAGLI; CLINIO; RAYCHTOCK, 2014) que engloba várias vertentes que se complementam, mas que por si só abrem leques de possibilidades. Esse movimento se apresenta como um grande desafio que impacta não somente na disseminação dos resultados em artigos bem como nas atividades intermediárias de pesquisa (cadernos de pesquisa entre outros), assim como reflete nos pesquisadores à medida que altera o processo de comunicação científica tradicional. Logo, as iniciativas em prol da CA buscam garantir mais transparência, possibilidades de acesso e uso do conhecimento científico.

Para tanto, a gestão dos dados de pesquisa é fundamental nesse processo já que é uma atividade que envolve a geração, o manuseio, o armazenamento e a descrição dos dados. “A gestão adequada de dados abertos pode apresentar-se como potencializadores para gerar o conhecimento científico, e como consequência favorecer avanços para a pesquisa científica e para a sociedade [...]” (RIBEIRO; OLIVEIRA; ARAÚJO, 2019, p. 163).

Nesse sentido, ao realizar a matriz de associação de Spearman (Quadro 12), observa-se que na categoria Ciência Aberta, houve uma associação quase nula ($r_s = -0,08$) entre as proposições (P1 e P2) e também que não houve significância na associação ($p\text{-valor} <$

0,47), logo não é possível inferir essa associação para além da amostra coletada.

Quadro 12 - Matriz de associação de Spearman entre as proposições da categoria “Ciência Aberta”

	P1	P2
P1	1,00	- 0,08
P2	- 0,08	1,00

Fonte: Dados da pesquisa.

A respeito do conjunto de proposições sobre gestão de dados de pesquisa, o sentimento prioritário foi de concordância. O GCp (51,81) apontou uma concordância desprezível, corroborado pelo Cp (172) que indicou que uma pequena maioria dos respondentes concordaram com a relevância da gestão de dados ao tempo que apontaram dificuldades e falta de apoio da instituição nesse processo, como ilustra a Quadro 13. Essa concordância evidencia lacunas e/ou falta de diretrizes que contribuem na gestão e isto é destacado na proposição de nº 5 com Cp de 66,5.

Quadro 13 – Gestão de dados de pesquisa

Proposição Categoria 2: Gestão de Dados de Pesquisa	Diferencial Semântico					T o t a l	Discordantes da proposição (D _p)	Concordantes da proposição (C _p)	Grau de Concordância da proposição (GC _p)
	1-DT	2-DP	3-I	4-CP	5-CT				
3. Preciso de fonte de dados confiáveis para o projeto de pesquisa.	0	0	2	0	81	83	82	1	51,81
4. A minha Instituição possui infraestrutura para a gestão dos meus dados	19	17	17	23	7	83	44,5	38,5	
5. Faltam diretrizes para a reutilização de dados na minha Instituição.	1	6	19	33	24	83	16,5	66,5	
6. Percebo dificuldades para implementação da política de gestão de dados científicos na minha Instituição.	2	9	12	38	22	83	17	66	
Total	160		172		332	0,4819	0,5181		
	D_p		C_p			μ2	μ1		

Fonte: Elaboração própria.

É consensual a relevância do uso dos dados de pesquisa científica no processo de criação da informação e avanço científico, bem como a qualidade e a proveniência deles no

compartilhamento (CUNHA; COSTA, 2020). Os dados abaixo demonstram uma realidade remota dos pesquisadores da UFBA em relação ao que tem acontecido mundialmente e em outras universidades do Brasil, como por exemplo, o “Manifesto de Acesso Aberto a Dados da Pesquisa Brasileira para a Ciência Cidadã” pelo Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict) que objetiva estimular o acesso livre, abrangente e ilimitado aos dados de pesquisa (fontes primárias), tornando possível o uso, reuso, compartilhamento e reproduzibilidade da pesquisa.

Vale trazer à tona também a pretensão da Horizon Europa 2020 em tornar públicas todas as publicações científicas resultantes de pesquisas financiadas pelo governo, exigindo que os cientistas gerenciem seus dados de pesquisa, elaborem Planos de Gestão de Dados e publiquem os resultados das pesquisas em periódicos de acesso aberto ou em plataformas de acesso compatíveis. Essa ação precisa ser incentivada pela Instituição de Ensino.

Ao analisar a perspectiva de Foster para uma gestão dos dados de pesquisa (Figura 5), a UFBA está muito aquém. Os pesquisadores apontaram uma necessidade de políticas que coadunem com o avanço da CA e que possibilitem a gestão dos dados de pesquisa científicas, mas cabe lembrar que gerir dados exige uma cooperação entre os atores envolvidos: pesquisadores, bibliotecas, instituições e agências financeiras, cada um com seu papel responsável por partes do ciclo de vida dos dados (GOLD, 2007). Falta na UFBA políticas que compreendam e pratiquem as mudanças reais na comunicação científica.

Em relação a esta categoria, observa-se que o valor-p dos pares de perguntas 4 e 5, 4 e 6, 5 e 6 foram significativas, indicando que há correlação entre elas. De acordo com o resultado do coeficiente de Spearman (Quadro 14), as relações significativas são positivas (P5 e P6) e negativas (P4 e P5, P4 e P6), sendo as mais fortes: positiva: P5 e P6 ($r_s = 0,61$), negativa: P4 e P5 ($r_s = -0,52$).

Quadro 14 - Matriz de associação de Spearman entre as proposições da categoria “Gestão de dados de pesquisa”

	P3	P4	P5	P6
P3	1,00	0,10	- 0,02	- 0,02
P4	0,10	1,00	- 0,52	- 0,47
P5	- 0,02	- 0,52	1,00	0,61
P6	- 0,02	- 0,47	0,61	1,00

Fonte: Dados da pesquisa.

A fim de conhecer a opinião dos pesquisadores sobre a categoria “Atitudes e valores em relação a abertura de dados”, as respostas dos questionados se dirigem ao preceito apontado pela literatura (COSTA, 2017) uma vez que a maioria dos questionados concordam que a abertura dos dados possibilita inúmeros benefícios: aumenta a credibilidade das pesquisas, fortalece o avanço científico, contribuem para a transparência, credibilidade e maior integridade da pesquisa, caracterizando um sentimento geral de concordância total (CT) quanto à assertiva e uma concordância muito forte é o que demonstra o GCP (94,82), ilustrado no Quadro 15.

As discussões em torno da abertura dos dados de pesquisas têm se manifestado como um marco importante na evolução dos projetos de Ciência Aberta que se estabelece a partir de conexões de elementos e atores envolvidos, em especial, as instituições produtoras de conhecimento científico, editores e agências de fomento.

Quadro 15 – Atitudes e valores em relação a abertura de dados

Proposição Categoria 3: Atitudes e valores em relação a abertura de dados	Diferencial Semântico					T o t a l	Discordantes da proposição (D _p)	Concordantes da proposição (C _p)	Grau de Concordância da proposição (GC _p)
	1-DT	2-DP	3-I	4-CP	5-CT				
7. Acredito que aumenta a credibilidade da minha pesquisa.	2	3	3	30	45	83	6,5	76,5	94,82
8. Contribui para o fortalecimento do avanço científico já que esses dados podem ser utilizados por outros pesquisadores em suas pesquisas.	0	2	2	18	61	83	3	80	
9. Contribui com a transparência na conduta da pesquisa.	1	1	3	20	58	83	3,5	79,5	
10. A abertura de dados de pesquisa contribui para uma maior integridade de na pesquisa	1	1	4	27	50	83	4	79	
11. Contribui para uma reproduzibilidade da pesquisa.	1	2	3	22	55	83	4,5	78,5	
Total	21,5		393,5			415	0,0518	0,9482	
	D _p		C _p				μ2	μ1	

Fonte: Elaboração própria.

A respeito da matriz de associação, todos os valores foram significantes e positivos, o que indica que as proposições possuem uma associação positiva entre elas (Quadro 16), sendo a mais forte a proposição P8 e P9 ($r_s = 0,79$).

Quadro 16 - Matriz de associação de Spearman entre as proposições da categoria “Atitudes e valores em relação a abertura de dados”

	P7	P8	P9	P10	P11
P7	1,00	0,55	0,60	0,50	0,55
P8	0,55	1,00	0,73	0,56	0,46
P9	0,60	0,73	1,00	0,56	0,53
P10	0,50	0,56	0,56	1,00	0,52
P11	0,55	0,46	0,53	0,52	1,00

Fonte: Dados da pesquisa.

Ao ser questionado sobre os motivos para abertura ou não dos dados científicos coletados pelo pesquisador durante a fase da pesquisa, os respondentes sinalizaram concordância em: acessibilidade, confiabilidade, reconhecimento profissional e visibilidade dos dados, conforme ratifica o Quadro 17. Das respostas que compõem concordantes das proposições, pode-se dizer que do total de respondentes, $C_p = 332,5$ concordaram com a proposição, de modo que $D_p = 82,5$ dos respondentes discordaram da proposição. Desta maneira, conforme Davis (1976), já exposto no Quadro 8, o valor da GC_p (80,12) indica que existe uma concordância substancial dos respondentes com relação a todas as proposições dessa categoria.

Quadro 17 – Motivação para a abertura

Proposição Categoria 4: Motivação para a abertura de dados	Diferencial Semântico					T o t a l	Discordantes da proposição (Dp)	Concordantes da proposição (Cp)	Grau de Concordância da proposição (GCp)
	1- DT	2- DP	3- I	4- CP	5- CT				
12. Amplia a acessibilidade da minha pesquisa, garantindo que o material seja acessível de maneira estável por longo tempo.	0	1	3	27	52	83	2,5	80,5	80,12
13. Possibilita maior confiabilidade já que o material em repositório tem credibilidade.	1	6	5	28	43	83	9,5	73,5	
14. Aumenta o prestígio em minha área e, consequentemente, reconhecimento profissional.	1	8	14	21	39	83	16	67	
15. Amplia a visibilidade da minha pesquisa.	0	2	6	27	48	83	5	78	
16. Sou obrigado(a) a abrir por causa da minha agência financiadora.	26	12	23	0	4	83	49,5	33,5	
Total	82,5			332,5		415	0,1988	0,8012	
	Dp			Cp			μ2	μ1	

Fonte: Elaboração própria.

É importante observar a proposição 16, esta obteve menor grau de concordância, com Dp equivalente a 49,5. Com base em Davis (1976), a interpretação desse valor apresenta uma discordância desprezível, o que mostra que a exigência da agência financiadora não seria o motivo para a abertura dos dados, mas sim os critérios acima elencados.

Segundo Grigg (2015), a abertura dos dados da pesquisa se dá por alguns princípios: a publicidade, o que possibilita maior visibilidade aos pesquisadores; o reuso de dados em novas conexões; a indução da colaboração em rede e a transparência dos dados utilizados para os resultados; a aceleração da produção de novas pesquisas; o atendimento às regras de financiadoras de pesquisa (que não é o caso, segundo aponta este estudo); o aperfeiçoamento dos investimentos em pesquisa e a redução da duplicação destes; a promoção da reproduzibilidade, da verificabilidade para garantia da boa prática científica; o acesso à pesquisa de importância social e uma maior consciência dos desafios da sociedade.

Considerando a relação entre governo aberto e o movimento da abertura dos dados de pesquisa, e que os dados produzidos nas atividades de pesquisa das instituições públicas (atividades governamentais), estes devem obedecer às normas legais e aos objetivos delineados pelo governo aberto. Porém, os pesquisadores precisam estar assistidos. O

movimento da Ciência Aberta envolve atores fundamentais no progresso das universidades que devem contribuir com criação de políticas com suas diretrizes que auxiliem os pesquisadores nesse avanço.

Sobre a matriz, os resultados do *p-valor* dos pares das proposições 12 e 13, 12 e 14, 12 e 15, 13 e 14, 13 e 15, 14 e 15 foram significativos (Quadro 18), indicando que eles possuem uma associação entre elas. Observando os valores do coeficiente de Spearman, tem-se que todos possuem uma relação positiva com relação às proposições 14 e 15 ($r_s = 0,81$) sendo essas as mais fortes.

Quadro 18 - Matriz de associação de Spearman entre as proposições da categoria “Motivação para a abertura”

	P12	P13	P14	P15	P16
P12	1,00	0,59	0,35	0,46	- 0,02
P13	0,59	1,00	0,60	0,64	0,07
P14	0,35	0,60	1,00	0,81	0,08
P15	0,46	0,64	0,81	1,00	0,11
P16	- 0,02	0,07	0,08	0,11	1,00

Fonte: Dados da pesquisa.

Quando perguntado a respeito das necessidades do pesquisador sobre gestão de dados de pesquisa científica na sua Instituição, a maioria sinalizou necessidades de apoio técnico na elaboração de PGD, de serviço de armazenamento e preservação dos dados, bem como de *backup* de arquivos e consultoria na gestão de dados o que inclui familiaridade com as ferramentas que auxiliam a gestão, com ênfase na tecnologia da informação. Autores como Gray (2007) e Hey e Trefethen (2003), observam que os desafios tecnológicos englobam a necessidade de melhor captar, analisar, visualizar e preservar as informações científicas, tornado a TI fundamental no atual ambiente de pesquisa, já que na Ciência Aberta, o encontro entre tecnologia e cientistas é inevitável. Como se vê no Quadro 19, em todas as proposições, a concordância foi prioritária e é afirmada pela GCP que corresponde a 90,51, evidenciando concordância muito forte.

É importante refletir sobre a necessidade apontadas pelos pesquisadores, pois a falta de conhecimento ou habilidade no tratamento adequado que possa viabilizar o processo de coleta, armazenamento, organização, busca, recuperação, difusão e preservação dos dados científicos primários coletados em grande escala, pode ocasionar uma perda em massa dos dados por falta de apoio técnico no tratamento adequado dos dados.

Quadro 19 – Necessidades dos pesquisadores

Proposição Categoria 5: Necessidades dos pesquisadores	Diferencial Semântico					T o t a l	Discordantes da proposição (D _p)	Concordantes da proposição (C _p)	Grau de Concordância da proposição (G _{Cp})
	1- DT	2- DP	3- I	4- CP	5- CT				
17. Apoio técnico na elaboração do plano de gestão de dados, ou seja, identificação de normas e esquema disciplinares de descrição de dados.	6	2	12	14	49	83	14	69	90,51
18. Necessito de serviço de armazenamento e preservação, do tipo <i>software</i> que possa contribuir no armazenamento e preservação dos meus dados.	1	1	9	10	62	83	6,5	76,5	
19. Necessito de <i>backup</i> de arquivos.	1	1	4	12	65	83	4	79	
20. Preciso de formação e/ou consultoria na gestão de dados de pesquisa	2	2	6	19	54	83	7	76	
Total	31,5			300,5		332	0,0949	0,9051	
	D _p			C _p			μ2	μ1	

Fonte: Elaboração própria.

Sobre a categoria, os resultados do *p*-valor dos pares das proposições 17 e 18, 17 e 20, 18 e 19, 18 e 20, 19 e 20 foram significativos, indicando que existem associações entre as proposições da categoria “Necessidade dos pesquisadores”. O valor do coeficiente de Spearman (Quadro 20), nos indica que as associações entre essas proposições são positivas, sendo a maior delas entre 17 e 18 ($r_s = 0,51$).

Quadro 20 - Matriz de associação de Spearman entre as proposições da categoria “Necessidades dos pesquisadores”

	P17	P18	P19	P20
P17	1,00	0,51	0,12	0,41
P18	0,51	1,00	0,41	0,42
P19	0,12	0,41	1,00	0,26
P20	0,41	0,42	0,26	1,00

Fonte: Dados da pesquisa.

As agências de fomento foi outro assunto abordado nessa pesquisa. O reconhecimento da relevância dos dados de pesquisa pelas agências na qual exigem uma adequada gestão dos dados científicos mostra uma necessidade de se ter mecanismos para preservação e reutilização desses dados. Esse mecanismo é feito por meio de políticas que regularizam a

abertura de dados de pesquisa oriundos de recursos públicos. No âmbito internacional, o United States National of Health (NIH), por exemplo, além de exigir que os artigos de todas as pesquisas financiadas sejam depositados no repositório institucional Pubmed Central, requer também que os dados de pesquisa sejam disponíveis em acesso aberto. Outro Exemplo, é o programa Horizon Europe 2020, tem como requisito para a participação e pagamento de bolsas a candidatos, a publicação em acesso aberto (EUROPEAN COMMISSION, 2020; GRIGG, 2015). No âmbito estadual, não há agência que exija o plano de gestão de dados científico. Isso explica o resultado apresentado no Quadro 21.

Em relação à agência de fomento, quando questionado sobre a exigência solicitada como condição para compartilhar, a concordância foi de 46,5, no entanto a maioria afirmou não saber elaborar o plano de gestão de dados. Nesse sentido, o GCP da categoria é equivalente a 56,63, o que diz respeito a uma concordância desprezível. Isso indica que as mudanças políticas das agências de fomento precisam refletir nas políticas institucionais das universidades e, consequentemente, atribuir ao pesquisador a publicação dos dados de pesquisa com o intuito de promover um ambiente social favorável à democratização da ciência. E as Universidades precisam criar maneiras de ajudar os pesquisadores na execução das políticas de acessibilidade, segundo a Open Science and Research Initiative (2014), essa é a principal premissa da pesquisa aberta.

Quadro 21 – Agência de fomento

Proposição Categoria 6: Agência de fomento	Diferencial Semântico					T o t a l	Discordantes da proposição (D_p)	Concordantes da proposição (C_p)	Grau de Concordância da proposição (G_Cp)
	1- DT	2- DP	3- I	4- CP	5- CT				
21. A agência de fomento exige um plano de gestão de dados no projeto de sua pesquisa, mas não sei elaborar esse documento.	14	12	19	24	14	83	35,5	47,5	56,63
22. Compartilho os dados para atender a demanda da agência de fomento.	15	9	25	20	14	83	36,5	46,5	
Total	72			94		166	0,4337	0,5663	
	Dp			Cp			μ2	μ1	

Fonte: Elaboração própria.

Com base nos dados do Quadro 22 há uma associação entre as proposições 21 e 22 sendo a mesma positiva ($r_s = 0,38$) e significativas ($p\text{-valor} > 0,000$).

Quadro 22 - Matriz de associação de Spearman entre as proposições da categoria “Agência de fomento”

	P21	P22
P21	1,00	0,38
P22	0,38	1,00

Fonte: Dados da pesquisa.

É sabido que a gestão de dados é um pilar da Ciência Aberta. Na gestão de dados, elementos distintos atribuem complexidade para a gestão e o compartilhamento dos dados da pesquisa, dentre eles Medeiros (2016) apontam a propriedade intelectual e o direito autoral. Cada vez mais evidente e discutido, a gestão de dados, dentre outras ações, deve respeitar a privacidade e segurança dos atores envolvidos no que tange a propriedade intelectual e as determinações legais.

Nesse sentido, ao ser questionado sobre direito autoral e proteção de dados (Quadro 23), os respondentes sinalizam uma concordância (Cp 227,5), enquanto que a discordância foi bem próxima Dp (187,5). No entanto, a GCp (54,82) interpreta como uma concordância desprezível. Ora, a propriedade intelectual é um dos motivos da “recusa” do compartilhamento de dados. Óbvio que em os critérios dos tipos de dados, mas o que prevalece é a obsessão enraizada devido a pressão nos ambientes acadêmicos e universitários “[...] para o patenteamento e o retorno financeiro dos resultados das atividades de Ciência e Tecnologia, mobilizando o estabelecimento de aparatos institucionais e legislação pertinentes a esses objetivos [...]” (ALBAGLI, 2015, p. 11). Outro fator importante, é a dependência das publicações científicas no que tange aos valores cobrados pelas assinaturas e a imposição restritivas ao acesso e uso.

Quadro 23 – Direito autoral/Proteção de dados

Proposição Categoria J : Direito autoral/Proteção de dados	Diferencial Semântico					T_{9,1}	Discordantes da proposição (D_p)	Concordantes da proposição (C_p)	Grau de Concordância da proposição (GC_p)
	1- DT	2- DP	3- I	4- CP	5- CT				
23. Meus dados são sensíveis (potencialmente sigilosos).	15	13	11	30	14	83	33,5	49,5	54,82
24. Meus dados são ostensivos (sem sigilo).	8	20	14	23	18	83	35	48	
25. Meus dados podem ou geram patentes.	42	7	14	13	7	83	56	27	
26. Meus dados são pessoais.	26	10	10	31	6	83	41	42	
27. Meus dados têm parcerias interinstitucionais.	14	1	14	33	21	83	22	61	
Total	187,5		227,5		Cp	415	0,4518	0,5482	
	Dp						μ2	μ1	

Fonte: Elaboração própria.

Conforme se pode observar nos dados do Quadro 23 e pelo coeficiente de Spearman (Quadro 24), entende-se que algumas relações são positivas, entre elas P23 e P25, P23 e P26, P25 e P27 e outras são negativas como P23 e P24, P24 e P27, sendo as mais fortes: positiva entre P23 e P26 ($r_s = 0,50$) negativa P23 e P24 ($r_s - 0,31$). Ressalta-se ainda que o p-valor dos pares de proposições 23 e 24, 23 e 25, 23 e 26, 24 e 27, 25 e 27 foram significativos.

Quadro 24 - Matriz de associação de Spearman entre as proposições da categoria “Direito autoral/Proteção de dados”

	P23	P24	P25	P26	P27
P23	1,00	- 0,31	0,30	0,50	0,20
P24	- 0,31	1,00	- 0,04	- 0,20	- 0,28
P25	0,30	- 0,04	1,00	0,04	0,35
P26	0,50	- 0,20	0,04	1,00	0,19
P27	0,20	- 0,28	0,35	0,19	1,00

Fonte: Dados da pesquisa.

A respeito da necessidade de um profissional capacitado e/ou biblioteca para o assessoramento, junto ao pesquisador, no processo de gestão dos dados de pesquisa, do total de respondentes, a concordância foi de Cp 210,5, enquanto que a discordância foi de Dp

(121,5). Esses dados trazem GCp (63,40), o que corresponde a uma concordância baixa (Quadro 25). Chama atenção a afirmação da necessidade de profissional capacitado para orientar em todo processo da gestão, mas é sutil a concordância dos pesquisadores respondentes em visualizar que a biblioteca tem esse papel também de contribuir nas orientações para o gerenciamento dos dados de pesquisa. Outro ponto é que, para os respondentes, a instituição não contribui para a formação de profissional capaz de realizar o gerenciamento de dados científicos.

Quadro 25 – Necessidades de profissional capacitado e/ou Biblioteca

Proposição Categoria 8: Necessidade profissional capacitado e/ou Biblioteca	Diferencial Semântico					Total	Discordantes da proposição (Dp)	Concordantes da proposição (Cp)	Grau de Concordância da proposição (GCp)
	1- DT	2- DP	3- I	4- CP	5- CT				
28. Sinto necessidade de profissional capacitado para orientar em todo processo da gestão dos meus dados científicos, desde a gênese até o armazenamento e preservação.	2	8	4	31	38	83	12	71	63,40
29. Percebo que a minha Instituição tem contribuído para a formação de profissional capaz de realizar o gerenciamento de dados científicos.	24	19	15	7	18	83	50,5	32,5	
30. A biblioteca poderia gerenciar os meus dados científicos por meio de um repositório de dados.	10	19	9	31	14	83	33,5	49,5	
31. Não é papel da Biblioteca gerenciar os meus dados e sim eu, pesquisador. Cabe uma Assessoria fornecida pela Pró-reitoria.	10	10	11	32	20	83	25,5	57,5	
Total	121,5			210,5		332	0,3660	0,6340	
	Dp			Cp			μ2	μ1	

Fonte: Elaboração própria.

Na verdade, os serviços de gestão de dados de pesquisa incluem: educação, expertise e curadoria e devem ser gerenciados pelas bibliotecas e requer trabalhos em parceria com setores da universidade. Segundo o relatório da OCLC (2017), o papel das bibliotecas na prestação de serviços de gestão de dados aos pesquisadores é essencial, pois os bibliotecários oferecem suporte à realização de revisões de fontes literárias, capacitam no uso dos recursos e fontes científicas de informação. Além disso, treinam os usuários para a utilização de gerenciadores de referências e citações, esclarecem sobre as políticas de acesso aberto, licenças e direitos autorais, instruem o depósito da produção intelectual em repositórios etc. Ou seja, são inúmeros os serviços de apoio aos pesquisadores, além de fornecer acesso a

dados de pesquisa, curadoria, armazenamento e preservação de dados de pesquisa.

Nos Estados Unidos, por exemplo, os serviços de gestão de dados de pesquisa são gerenciados pelas bibliotecas em conjunto com a equipe de informática e com as pró-reitorias, é o caso da Universidade of Illinoi at Urbana Chmapain, da Monash University, na Austrália, entre outros. Mesmo assim, há quem entenda que o bibliotecário e a biblioteca devam se limitar apenas às suas atividades técnicas. Uma visão arcaica que aos poucos está sendo transformada.

Cabe destacar que, é latente a atuação do bibliotecário e das bibliotecas, contudo o movimento da CA reverbera a necessidade de atualização destes profissionais, principalmente nas técnicas de preservação e conservação, com ênfase para os materiais especiais e as técnicas de bibliometria, pois é importante a avaliação quanto a obsolescência do acervo, legal e de metodologias científicas para suportar todos o gerenciamento informacional (FIGUEIREDO, 1979, 1998).

No que diz respeito à matriz de Spearman, observa-se que o p-valor dos pares de proposições 28 e 30, 29 e 30, 29 e 31, 30 e 31 são significantes, indicando que há associação entre elas. Pelo resultado do coeficiente de Spearman (Quadro 26), há relações positivas (P28 e P30, P29 e P30) e negativas (P29 e P31, P30 e P31), sendo as mais fortes: positiva, as proposições P29 e P30 ($r_s = 0,30$); negativa, as proposições P30 e P31 ($r_s = -0,40$).

Quadro 26 - Matriz de associação de Spearman entre as proposições da categoria “Necessidades de um profissional capacitado e/ou biblioteca”

	P28	P29	P30	P31
P28	1,00	0,01	0,22	0,08
P29	0,01	1,00	0,30	-0,25
P30	0,22	0,30	1,00	-0,40
P31	0,08	-0,25	-0,40	1,00

Fonte: Dados da pesquisa.

Sobre as razões para não compartilhar dados de pesquisa, prevaleceram a falta de conhecimento e o medo de haver uma má interpretação do uso e a usurpação dos dados. Essa concordância foi equivalente a GCp (64,61), uma concordância baixa (Quadro 27). Essa preocupação por parte dos pesquisadores é apontada no relatório “Open data: unlocking innovation and performance with liquid information” realizado pela McKinsey. O relatório infere que, talvez, a publicação de dados de pesquisa não seja uma das prioridades mais

premente dos pesquisadores (MANYIKA *et al.*, 2013).

Justificativas para não compartilhar é o que não faltam, é o que aponta um estudo coordenado pela médica Livia Puljak, da Universidade Católica da Croácia, e publicado na *Revista Pesquisa Fapesb*, em 11 de julho de 2022. Para ela, “parece haver um abismo entre o que os pesquisadores dizem e o fazem”⁵³. Nesse estudo, Puljak e colaboradores (2022) mostram que, mesmo aqueles cientistas que declaram e se comprometem a compartilhar seus dados, não os fazem na prática e dão inúmeras justificativas para negação: abandono de projeto, perdas de dados ou até mesmo que não tem autorização para compartilhar alguns desses dados.

Quadro 27 – Razões para não compartilhar dados de pesquisa

Proposição Categoria 9: Razões para não compartilhar dados de pesquisa	Diferencial Semântico					T o t a l	Discordantes da proposição (D_p)	Concordantes da proposição (C_p)	Grau de Concordância da proposição (GC_p)
	1-DT	2-DP	3-I	4-CP	5-CT				
32. Não sei como compartilhar dados. Preciso de suporte (assessoria).	8	10	5	30	30	83	20,5	62,5	64,61
33. Meus dados não podem estar abertos, pois contêm informação sigilosa.	18	14	13	30	8	83	38,5	44,5	
34. Mau uso dos dados pode causar uma má interpretação por outros pesquisadores.	7	6	11	26	33	83	18,5	64,5	
35. Se eu compartilhar, meus dados podem ser usurados por outros pesquisadores.	16	19	10	31	7	83	40	43	
Total	117,5			214,5		332	0,3539	0,6461	
	D_p			C_p			μ2	μ1	

Fonte: Elaboração própria

Observa-se que o p-valor dos pares de perguntas 32 e 33, 32 e 35, 33 e 34, 33 e 35, 34 e 35 são significantes, indicando que há associações entre elas. Pelo coeficiente de Spearman (Quadro 28), tem-se que as relações são positivas, com a mais forte sendo a relação entre P34 e P35 ($r_s = 0,37$).

⁵³ Disponível em: Cientistas ainda resistem em compartilhar dados de pesquisa: *Revista Pesquisa Fapesp*.

Quadro 28 - Matriz de associação de Spearman entre as proposições da categoria “Razões para não compartilhar dados de pesquisa”

	P32	P33	P34	P35
P32	1,00	0,31	0,17	0,36
P33	0,31	1,00	0,27	0,26
P34	0,17	0,27	1,00	0,37
P35	0,36	0,26	0,37	1,00

Fonte: Dados de pesquisa

Uma das facetas da CA é o repositório de dados. O movimento de compartilhamento exige a necessidade de se ter um repositório de dados tratados para se ter a recuperação e que se possa difundir a informação e ser usada em outras pesquisas. Para Monteiro (2021), os repositórios de dados possibilitam o armazenamento e a gestão de dados vislumbrando o aprimoramento da recuperação, aumentando as potencialidades de reuso dos dados entre os pesquisadores e otimiza os processos da investigação científica, resultando no avanço da ciência.

A respeito disso, quando questionados sobre repositório de dados (Quadro 29), chama atenção às discordâncias apontadas pelos respondentes Dp (266,5), enquanto que a concordância foi de Cp (231,5). Desta maneira obteve como GCp (46,49), o que na concepção de Davis (1976), é uma discordância desprezível. Isso indica baixa adesão no uso de repositório, o que mostra pouco uso deles tanto para compartilhar quanto para uso de dados de outros pesquisadores. O desconhecimento de repositório de dados é um dos fatores que mais influencia o pesquisador a não compartilhar seus artigos científicos, afirma Rodrigues (2014). Porém, não é possível saber se o cenário seria diferente caso eles tivessem conhecimento a respeito, já que recompensa acadêmica e promoção na carreira ainda são latentes.

Sobre isso, o projeto Rede de Dados de Pesquisa Brasileira criado pelo o Grupo de Trabalho da Rede de Dados de Pesquisa Brasileira (RDP Brasil)⁵⁴, tem como objetivo 1, identificar as práticas de AADP em instituições brasileiras. Diante disso, foi realizado um levantamento no Re3Data – registro global de repositórios de dados de pesquisa que abrange

⁵⁴ A Rede Nacional de Ensino e Pesquisa (RNP) e o Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict) em colaboração com a Universidade Federal do Rio Grande do Sul (UFRGS) e Universidade do Rio Grande (FURG) formaram o Grupo de Trabalho da Rede de Dados de Pesquisa Brasileira (RDP Brasil) para o desenvolvimento de um projeto em conjunto sobre o Acesso Aberto a Dados de Pesquisa (AADP). O objetivo do Grupo de Trabalho da Rede de Dados de Pesquisa Brasileira (RDP Brasil) é executar atividades que contribuam para a identificação de práticas, mapeamento de requisitos e prototipação de sistemas que facilitem a disseminação de informações científicas. O projeto visa colaborar com a criação de um Programa Nacional de Acesso Aberto a Dados de Pesquisa, uma iniciativa com o objetivo de promover e incentivar o compartilhamento de dados entre pesquisadores, conferindo maior eficiência na produção de conhecimento científico no Brasil. Para mais informações ver: RDP Brasil (ufrgs.br).

diferentes áreas do conhecimento – a fim de identificar repositórios de Acesso Aberto a Dados de Pesquisa (AADP) em universidades e em instituições de pesquisa brasileiras. Os resultados mostraram uma diversidade de repositórios de dados no Brasil e foram citados no relatório do RDPBrasil 2018, a saber:

- a) Ibict DataVERSE Network⁵⁵ disponibiliza um repositório de dados de pesquisa que cuida da preservação a longo prazo e de boas práticas de arquivamento, para que os pesquisadores possam compartilhar, manter o controle e obter o reconhecimento de seus dados;
- b) Banco de Dados de Exploração e Produção (BDEP)⁵⁶ desde 2000 possui em seu acervo mais de seis petabytes de dados técnicos gerados nas atividades de prospecção petrolífera em todo território brasileiro;
- c) Global Collaboration Engine (GLOBE)⁵⁷ disponibiliza um ambiente colaborativo on-line que permite aos pesquisadores compartilhar, comparar e integrar estudos locais e regionais com dados globais para avaliar a relevância global de seu trabalho sobre mudança de uso e cobertura de solo;
- d) International Ocean Discovery Program (IODP)⁵⁸ é um programa de colaboração internacional de pesquisa marinha que explora a história e a dinâmica da Terra usando plataformas de pesquisa oceânica para recuperar dados registrados em sedimentos e rochas do fundo do mar e monitorar ambientes de subsolos;
- e) PPBio Data Repository⁵⁹ tem o objetivo de fomentar estudos sobre biodiversidade no Brasil, descentralizando a produção científica de centros acadêmicos já desenvolvidos, integrando atividades de pesquisa e disseminando resultados em diversas finalidades, incluindo gestão ambiental e educação;
- f) Global Climate Data (WorldClim)⁶⁰ é um conjunto de dados climáticos globais (grades de clima) com uma resolução espacial de cerca de um quilômetro quadrado. Os dados podem ser usados para mapeamento e modelagem espacial em um SIG ou com outros programas de computador;

⁵⁵Disponível em: <https://repositoriopesquisas.ibict.br>

⁵⁶ Disponível em:

<http://www.anp.gov.br/exploracao-e-producao-de-oleo-e-gas/dados-tecnicos/acesso-aos-dados-tecnicos>

⁵⁷ Disponível em: <http://globe.umbc.edu>

⁵⁸ Disponível em: <http://www.iodp.org>

⁵⁹ Disponível em: <https://ppbiodata.inpa.gov.br/metacatui/>

⁶⁰ Disponível em: <http://worldclim.org>

g) Base de Dados Científicos da Universidade Federal do Paraná⁶¹ reuni os dados científicos utilizados nas pesquisas que foram publicadas pela comunidade da UFPR em teses, dissertações, artigos de periódicos e outros materiais bibliográficos. Estes repositórios são mantidos ou têm participação de instituições brasileiras (PAVÃO *et al.*, 2018).

Além destes tem-se: Repositório de Dados do Programa de Pesquisa de Biodiversidade da Amazônia Ocidental (PPBIO); Repositório de Dados do Programa de Pesquisas Ecológicas de Longa Duração (PELD), Portal GEOINFO de infraestrutura de dados espaciais da EMBRAPA, o Portal da Biodiversidade (SISBio), entre outros.

Quadro 29 – Repositório de dados

Proposição Categoria 10: Repositório de dados	Diferencial Semântico					I o t a l	Discordantes da proposição (D_p)	Concordantes da proposição (C_p)	Grau de Concordância da proposição (G_{Cp})
	1- DT	2- DP	3- I	4- CP	5- CT				
36. Não confio na transparência e na autenticidade dos dados brutos em repositório de dados	21	26	19	16	1	83	56,5	26,5	46,49
37. Não tenho motivos/vantagens que incentivam a publicação de dados em um repositório de acesso aberto.	21	25	16	16	5	83	54	29	
38. Necessito de serviços de compartilhamento e acesso a dados.	1	2	20	33	27	83	13	70	
39. Não tenho conhecimento de nenhum repositório para fazer o compartilhamento.	19	13	7	20	24	83	35,5	47,5	
40. Não compartilho porque meus dados têm restrição contratual.	39	14	18	8	4	83	62	21	
41. Não compartilho porque posso perder uma boa publicação, caso eu compartilhe os dados.	23	18	9	27	6	83	45,5	37,5	
Total	266,5			231,5		498	0,5351	0,4649	
	D_p			C_p			μ2	μ1	

Fonte: Elaboração própria

Referente a associação de Spearman, observa-se no Quadro 30 que o p-valor dos pares das proposições 36 e 37, 36 e 40, 36 e 41, 37 e 41, 39 e 41, 40 e 41, são significativas, indicando que há correlação entre elas. Pelo resultado do coeficiente de Spearman, é possível observar que as relações significativas são positivas, sendo a mais forte entre as proposições P40 e P41 ($r_s = 0,46$).

⁶¹ Disponível em: <https://bdc.c3sl.ufpr.br/>

Quadro 30 - Matriz de associação de Spearman entre as proposições da categoria “Repositório de dados”

	P36	P37	P38	P39	P40	P41
P36	1,00	0,40	- 0,07	0,15	0,28	0,39
P37	0,40	1,00	- 0,19	0,17	0,21	0,41
P38	- 0,07	- 0,19	1,00	0,05	0,03	- 0,19
P39	0,15	0,17	0,05	1,00	0,05	0,37
P40	0,28	0,21	0,03	0,05	1,00	0,46
P41	0,39	0,41	- 0,19	0,37	0,46	1,00

Fonte: Dados da pesquisa.

Swan (2005) e Rodrigues (2014) mostram a influência da idade do pesquisador na opção de compartilhar dados de pesquisa. Para eles, pesquisadores seniores tendem a resistir ao uso da tecnologia, o que implicaria no compartilhamento ou não dos dados. A presente pesquisa contou com pesquisadores de faixa etária distinta, mas a maioria dos respondentes encontra-se na faixa etária entre 51 anos ou mais, o que talvez justifique o “desconhecimento” de gestão de dados e de repositórios específicos para esta finalidade, bem como o pouco compartilhamento dos dados de pesquisa.

Para compreender os principais obstáculos, assim como as perspectivas dos professores sobre o movimento de acesso aberto e, principalmente, gestão de dados de pesquisa científica na UFBA, se fez necessário sintetizar os resultados das proposições do questionário com base no grau de concordância da população com as categorias selecionadas, conforme ilustra o Quadro 31.

Quadro 31 - Síntese dos graus de concordância com base na lógica paraconsistente

Categorias	GCp	Frase adequada
1. Ciência Aberta	63,25	Concordância baixa
2. Gestão de dados de pesquisa	51,81	Concordância desprezível
3. Atitude e valores para abertura de dados	94,82	Concordância muito forte
4. Motivação para a abertura de dado	80,12	Concordância substancial
5. Necessidade dos pesquisadores	90,51	Concordância muito forte
6. Agência de fomento	56,63	Concordância desprezível
7. Direito autoral/ proteção de dados	54,82	Concordância desprezível
8. Necessidade profissional capacitado e/ou Biblioteca	63,40	Concordância baixa
9. Razões para não compartilhar dados de pesquisa	64,61	Concordância baixa
10. Repositório de dados	46,49	Discordância desprezível

Fonte: Elaboração própria.

Os principais obstáculos estão relacionados ao “desconhecimento do movimento da Ciência Aberta e de repositórios de dados”, “preocupação com os direitos autorais/proteção de dados”, a “necessidade de ter capacitação e/ou profissionais capacitados para auxiliar no compartilhamento” e principalmente falta de diretrizes que direcionem como fazer a gestão de dados de pesquisa. Chama atenção também “a exigência da agência de fomento”, mas influencia pouco ou quase nada no comportamento de compartilhamento de artigo científico em repositórios já que ainda na Bahia não há essa exigência via financiador.

Além dos dados do questionário, alguns professores apresentaram suas percepções a respeito dos assuntos abordados nos pesquisas. Com o intuito de exemplificar a preocupação desses pesquisadores da Universidade Federal da Bahia com a temática deste estudo, a seguir serão descritos alguns desses trechos. Um dos respondentes demonstrou preocupação com a falta de conhecimento sobre a política (gestão de dados de pesquisa) e ênfase na necessidade da instituição apresentar à comunidade as políticas de gestão de dados de pesquisa científica bem como de acesso aberto:

No geral, tenho a impressão que conheço muito pouco sobre esta política. É possível que a UFBA até tenha algum tipo de serviço que dê suporte e orientação sobre isso, mas desconheço. Acredito que como pesquisadora de humanas, o compartilhamento de dados precisa ser pensado de maneira específica, garantindo sigilo dos colaboradores e cuidados para a sua identificação. Talvez fosse importante também pensar em um período de

embargo para permitir o tempo de publicação de trabalhos originais. (*Respondente 1*).

Alinhado a isso, outro respondente diz o seguinte: “a UFBA deveria questionar a CAPES sobre o pagamento por acesso a periódicos científicos cujo conteúdo foi gerado com recursos públicos e deveriam ser de acesso aberto. E incentivar os professores a preferir publicação em revista de acesso aberto” (*Respondente 2*), ou seja, os participantes da pesquisa entendem que é importante que a UFBA dê um direcionamento sobre a gestão de dados científicos, criando diretrizes para o compartilhamento de dados.

Infere-se, com isso, que o diálogo com o pesquisador na Universidade Federal da Bahia é inexistente no que se refere ao avanço e a inserção da participação desse ator no movimento da Ciência Aberta, em especial, na abertura e compartilhamento dos dados de pesquisa, ou seja, inversamente proporcional ao que aponta a literatura sobre o avanço do acesso aberto no Brasil e no mundo. A UFBA precisa entender e chamar o principal interessado, o pesquisador, para o conhecimento amplo, da relevância de uma boa gestão de dados de pesquisa e do acesso aberto dos dados. Agora, o auxílio do uso e da existência de repositório, tornando-o essa tarefa mais simplificada, é papel do profissional da informação.

6 CONSIDERAÇÕES FINAIS

A produção exponencial de dados tem mudado a dinâmica no fluxo da comunicação científica no que se refere à coleta, armazenamento, recuperação, acesso e compartilhamento. Os dados de pesquisas científicas em si só fazem parte do processo de produção científica, contudo, as questões inerentes à gestão deles, no Brasil, ainda são tão simplificadas. A proposta do movimento Ciência Aberta, em tese, é justamente simplificar a forma como os dados são tratados, padronizados, preservados e compartilhados de maneira gratuita.

A respeito do objeto de estudo desta pesquisa, no âmbito internacional, a literatura abordada demonstrou avanços em termos de política de gestão de dados de pesquisa científica no Reino Unido, Estados Unidos, Austrália, Portugal, entre outros. No âmbito nacional, algumas universidades já têm demonstrado práticas que revelam a necessidade de uma boa gestão de dados de pesquisa, descrevendo as diretrizes e direcionando os pesquisadores à construção de plano de dados de seus projetos de pesquisa: USP, UFPR, UFSCar, assim como a Fapesp. Em contrapartida, a Universidade Federal da Bahia, é ausente na participação em iniciativas que direcionem o acesso e a gestão de dados de pesquisa oriundos de financiamento público. De fato, a temática de gestão de dados de pesquisa é contemporânea e tem sido, a cada dia, integrada no âmbito das universidades, uma vez que estas são as grandes protagonistas na produção, uso e disseminação do conhecimento.

É verdade que o movimento da Ciência Aberta tem sido discutido amplamente, mas percebe-se que falta convidar o pesquisador para a discussão. Isso porque, a pesquisa em questão mostrou que os pesquisadores têm pouco ou nenhum conhecimento sobre os temas inerentes apontados nesse estudo: gestão de dados de pesquisa, repositório de dados, Ciência Aberta etc. Embora reconheça a relevância da abertura e/ou compartilhamento, aponta-se obstáculos para a prática, quer seja falta de diretrizes estipuladas pela UFBA, ou preocupação com a propriedade intelectual ou até mesmo necessidade de profissionais que auxiliem nas formas de fazer um plano e/ou acesso a repositório de dados. Isso evidencia a necessidade do profissional da informação trabalhar em conjunto com os pesquisadores de cada área de conhecimento para dirimir a complexidade vista pelos cientistas na gestão dos seus dados.

As poucas iniciativas da UFBA no intuito de tornar a informação aberta, como o repositório institucional e a criação do plano de dado abertos sinalizam positividade no diálogo, mas demonstra vagarosidade na participação do desenvolvimento de pesquisa de ponta e no compartilhamento dos dados científicos entre os cientistas, seja em nível nacional

e/ou internacional. Frisa-se, com isso, a necessidade de uma política que norteie as ações dos pesquisadores quanto ao gerenciamento de seus dados de pesquisa, com plataformas que facilite o armazenamento e o compartilhamento de pesquisas em banco de dados, com a devida autoria creditada, a fim de proporcionar maior adesão e segurança dos pesquisadores nesse processo e auxílio em inovações.

Diante o exposto, recomenda-se: a criação de diretrizes para a gestão de dados de pesquisa na UFBA para nortear o gerenciamento desses dados; a criação de repositório de dados, em conjunto com pesquisador e profissional da informação adotando as diretrizes internacionais, como por exemplo, a OpenAIRE Guidelines for Data que aborda recomendações o uso pelos repositórios de formas de citação de dado de pesquisa (*datacite*) com o intuito de minimizar as desconfianças em relação ao plágio; políticas claras de uso de dados compartilhados a fim de dirimir as preocupações sobre a propriedade intelectual; eventos, debates mais específicos sobre o movimento da Ciência Aberta, gestão de dados de pesquisa com objetivo de ampliar o conhecimento sobre essas temáticas e os benefícios para a evolução da ciência; cursos de capacitação para os usuários pesquisadores e os profissionais envolvidos nos processos que englobam o ciclo de vida dos dados, em especial o armazenamento e a preservação deles.

Sugere-se também um diálogo com os editores científicos e revistas da universidade, embora não evidenciado neste estudo até porque não foi esse o objeto, mas entende-se que essa conversa tangencia decisões importantes para mudanças na forma de se comunicar as pesquisas científicas. O objetivo seria analisar as políticas imprescindíveis dos editores científicos, averiguando as influências no índice de autoarquivamento dos repositórios, pois essas políticas influenciam no compartilhamento ou não dos dados. Talvez esse ponto seja mais interessante e outro estudo.

Como resultado dessa pesquisa, sugere-se que um estudo seja feito a fim de identificar o motivo ou as dificuldades, por parte da UFBA, da não implementação de normas para a gestão dos dados científicos. Outro ponto de investigação, é conhecer os motivos da agência de fomento na Bahia, Fapesb, não ter implementado uma política de dados científicos. É no mínimo curioso e preocupante, uma vez que tanto a universidade quanto a agência de fomento precisam se envolver nesse processo de progressão da ciência, definindo os papéis de cada um.

Como exposto, comprehende-se que os objetivos delineados para a condução teórica e

metodológica desta investigação foram atingidos, a saber: verificar na literatura os principais conceitos da Ciência Aberta, dados de pesquisa, plano de gestão de dados e FAIR; mapear as ações desenvolvidas para a implantação de uma política de gestão de dados de pesquisa na UFBA; identificar a percepção dos pesquisadores a respeito dos fatores que podem influenciar a abertura e/ou compartilhamento de dados de pesquisa científica; descrever os principais desafios e as soluções inerentes à construção de uma política estruturada para a gestão de dados de pesquisa científica na UFBA.

Para a análise da percepção dos pesquisadores sobre a temática em estudo, a utilização do questionário foi fundamental para mensurar o seu olhar em relação a cada assunto abordado que interligam e compreendem os conceitos e os princípios da CA e de gerenciamento de dados de pesquisa. Além disso, pesquisas anteriores documentadas na literatura científica e apresentadas no referencial teórico também foram primordiais para aprofundar as ações desenvolvidas na pesquisa.

Contudo, na busca para responder aos questionamentos apontados no estudo, ocorreram algumas limitações na coleta: a) os *sites* dos cursos de pós-graduação encontravam-se desorganizados, ou seja, nem sempre era possível diferenciar os professores permanentes, dos colaboradores e/ou visitantes. E quando tinham, alguns não constavam o contato que impossibilitou a aplicação do instrumento para o público desejado; b) as coordenações não respondiam às solicitações enviadas, ou seja, em virtude da falta de informações precisas, foi necessário solicitá-las às coordenações, mas nem sempre obtivemos respostas. Diante disso, foi necessário restringir esses professores dos públicos-alvo; c) falta de colaboração dos professores pesquisadores. Ou melhor, no decurso da aplicação do questionário, percebeu-se que a maioria dos professores permanentes não é muito adepto a participar de pesquisas que não sejam de seu interesse ou que não façam parte delas. Mesmo assim, os que contribuíram foram importantes para compreender as práticas e necessidades do sujeito e inferir algumas conclusões.

Diante disso, os resultados possibilitaram elaborar um projeto de plano de gestão de dados científicos na Universidade Federal da Bahia (APÊNDICE C). Espera-se contribuir para a implementação definitiva das normas para este fim na Instituição objeto desta pesquisa, e quiçá possibilitar o estado da arte do movimento da Ciência Aberta, em especial, da disponibilização de dados. É preciso incentivar e aprofundar o diálogo sobre os processos de comunicação científica que envolvem os dados de pesquisa, pois é notório que vários pontos

de inflexão ainda precisam ser compreendidos; por outro lado, é compreensível a progressão do movimento da CA está em fase de progressão o que justifica a busca constante por aprendizado sob suas diversas facetas.

Nesse viés, a Ciência da Informação, enquanto ciência interdisciplinar, preocupada com os fluxos informacionais, precisa discutir mais sobre as transformações da comunicação científica, nesse contexto de colaboração e compartilhamento de dados. Tais discussões resultariam em estratégias de promoção e divulgação científica, e quiçá formas de avaliação da eficácia desse movimento. É preciso, portanto, aprofundar nas discussões a respeito da temática

Acredita-se que as possíveis intervenções realizadas pela Instituição possam dirimir os obstáculos e somatizar estímulos imprescindíveis para a efetiva inserção do pesquisador ao conhecimento aberto já que o compartilhamento dos dados é parte essencial para a consolidação dos objetivos da Ciência Aberta. Logo, este estudo busca preencher as lacunas de conhecimento do campo e enriquece a discussão.

REFERÊNCIAS

- AALBERSBERG, I. J.; DUNHAM, J.; KOERS, H. Conectando artigos científicos com dados de pesquisa: novas direções na publicação acadêmica online. **Data Science Journal**, v. 12, p. WDS235-WDS242, 2013. DOI: <https://doi.org/10.2481/dsj.wds-043>. Disponível em: <https://datascience.codata.org/articles/abstract/10.2481/dsj.WDS-043/>. Acesso em: 14 ago. 2021.
- ABRAMS, S. *et al.* DataShare: capacitando a curadoria de dados dos pesquisadores. **International Journal of Digital Curation**, v. 9, n. 1, p. 110-118, 17 jun. 2014. DOI: <https://doi.org/10.2218/ijdc.v9i1.305>. Disponível em: <http://www.ijdc.net/article/view/9.1.110>. Acesso em: 14 ago. 2021.
- AKERS, K. G.; SFERDEAN, F. C.; NICHOLLS, N. H.; GREEN, J. A. Building Support for Research Data Management: Biographies of Eight Research Universities. **International Journal of Digital Curation**, v. 9, n. 2 p. 171-191, 2014. www.ijdc.net/index.php/ijdc/article/view/9.2.171/376 . Acesso em: 15 ago. 2021.
- ALBAGLI, S. Ciência aberta em questão. In: ALBAGLI, S.; MACIEL, M. L.; HANNUD ABDO, A. (org.). **Ciência aberta:** questões abertas. Brasília: IBICT; Rio de Janeiro: UNIRIO, 2015. p. 9-26. Disponível em: [http://livroaberto.ibict.br/bitstream/1/1060/1/Ciencia%20aberta_questoes%20abertas_PORTUGUES_DIGITAL%20\(5\).pdf](http://livroaberto.ibict.br/bitstream/1/1060/1/Ciencia%20aberta_questoes%20abertas_PORTUGUES_DIGITAL%20(5).pdf). Acesso em: 16 nov. 2020.
- ALBAGLI, S.; APPEL, A. L.; MACIEL, M. L. E-Science e ciência aberta: questões em debate. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 14., 2013, Florianópolis. **Anais...** Florianópolis: UFSC, 2013. Disponível em: <http://enancib.ibict.br/index.php/enancib/xivenancib/paper/viewFile/4242/3365>. Acesso em: 14 maio 2018.
- ALBAGLI, S.; CLINIO, A.; RAYCHTOCK, S. Ciência Aberta: correntes interpretativas e tipos de ação. **Liinc em Revista**, Rio de Janeiro, v. 10, n. 2, p. 434-450, nov. 2014. Disponível em: <http://revista.ibict.br/liinc/article/view/3593>. Acesso em: 22 out. 2020.
- ALBUQUERQUE, M. R. **A curadoria em um repositório institucional:** uma análise sob a ótica do ciclo de vida dos dados de Sant'ana. 2018. 87 f. Dissertação (Mestre em Ciência da Informação) - Universidade Federal do Ceará, Fortaleza, 2018.
- ALEXANDRE, N. M. C.; COLUCI, M. Z. O. Validade de conteúdo nos processos de construção e adaptação de instrumentos de medidas. **Ciência e Saúde Coletiva**, v. 16, n. 7, p. 3061-3068, 2011.
- ALMEIDA, F. G. **Supporte à gestão de dados de pesquisa :** uma ampliação dos serviços oferecidos pelas bibliotecas. 2019. 249 f. Tese (doutorado em gestão e organização do conhecimento) – Universidade Federal de Minas Gerais, Escola de Ciência da Informação, Belo Horizonte, 2019. Disponível em: https://repositorio.ufmg.br/bitstream/1843/31596/4/Tese_PPGGOC_Fernanda%20Gomes%20Almeida_11out.pdf. Acesso em: 12 ago. 2021.
- BAL, L. Guest Post: comunicação acadêmica aberta e mais rápida em um mundo pós-COVID. **The Scholarly Kitchen**, 23 jun. 2021. Disponível em: <https://scholarlykitchen.sspnet.org/2021/06/23/guest-post-open-and-faster-scholarly-communication-in-a-post-covid-world/>. Acesso em: 14 ago. 2021.
- BARBETTA, P. A. **Estatística aplicada às Ciências Sociais**. 6 ed. Florianópolis, SC:

- EDUFSC, 2006. Cap. 3.
- BARDIN, L. **Análise de conteúdo**. São Paulo: Edições 70, 2011.
- BARBOSA NETO, D. C. M.; ATAÍDE, J. A. R.; ZANCAN, C. Análise de redes sociais na produção científica de programas de pós-graduação em administração pública no Brasil. In: ENCONTRO BRASILEIRO DE ADMINISTRAÇÃO PÚBLICA, 6., João Pessoa, PB. **Anais...** João Pessoa, PB: UFPB, 2017.
- BERTIN, P. R. B. et al. A parceria para Governo Aberto como plataforma para o avanço da Ciência Aberta no Brasil. **Transinformação**, Campinas, v. 31, epub, 16 set. 2019a. DOI: <http://dx.doi.org/10.1590/2318-0889201931e190020>. Disponível em: <https://ainfo.cnptia.embrapa.br/digital/bitstream/item/212937/1/A-parceria-para-governo-aberto.pdf>. Acesso em: 14 fev. 2021.
- BERTIN, P. R. B. et al. A política de governança de dados, informação e conhecimento da Embrapa como mecanismo para a gestão de dados de pesquisa agropecuários. **Liinc em revista**, v. 15, n. 2, p. 194-204, 2019b. Disponível em: revista.ibict.br/liinc/article/view/4798. Acesso em: 12 set. 2019.
- BJÖRK, B-C. Open access to scientific publications: an analysis of the barriers to change. **Information Research**, v. 9, n. 2, paper 170, 2004. Disponível em: <http://InformationR.net/ir/9-2/paper170.html>. Acesso: 04 fev. 2021.
- BIAZON, T.; MARIN, T. Ciência aberta: uma nova forma de fazer ciência. **Com Ciência**, 10 maio 2016. Disponível em: <https://www.comciencia.br/comciencia/handler.php?section=8&edicao=123&id=1496>. Acesso em: 28 fev. 2021.
- BORKO, H. Information Science: What is it? **American Documentation**, v. 19, n. 1, p. 3-5, jan. 1968. (Tradução Livre). Disponível em: https://edisciplinas.usp.br/pluginfile.php/2532327/mod_resource/content/1/Oque%C3%A9CI.pdf. Acesso em: 12 out. 2019.
- 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: <http://www2.camara.gov.br>. Acesso em: 23 jan. 2020.
- BRASIL. Lei nº 10.973, de 2 de dezembro de 2004. Dispõe sobre incentivos à inovação e à pesquisa científica e tecnológica no ambiente produtivo e dá outras providências. **Diário Oficial da União**, Brasília, 2 dez. 2004. Disponível em: <https://www2.camara.leg.br/legin/fed/lei/2004/lei-10973-2-dezembro-2004-534975-norma-legalizada-pl.pdf>. Acesso em: 11 out. 2019.
- BRASIL. Lei nº 12.527 de 18 de novembro de 2011. 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**, Brasília, 18 nov. 2011. Disponível em: [<http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/l12527.htm>](http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/l12527.htm). Acesso em: 11 out. 2019.
- 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, 12 maio 2016. Disponível em: http://www.planalto.gov.br/CCIVIL_03/_Ato2015-2018/2016/Decreto/D8777.htm. Acesso

em: 11 out. 2019.

BORGES, M. A. G. A informação e o conhecimento como insumo ao processo de desenvolvimento. **Revista Ibero-americana de Ciência da Informação (RICI)**, v.1, n. 1, p.175-196, jul./dez. 2008. Disponível em: <http://164.41.122.25/portalnesp/ojs-2.1.1/index.php/rici>. Acesso em: 31 maio 2023.

BOURDIEU, P. **Usos sociais da ciência:** por uma sociologia clínica do campo científico. São Paulo: UNESP, 2004.

BOURDIEU, P. O campo científico. In: ORTIZ, Renato (org.). **Pierre Bourdieu:** sociologia. São Paulo: Ática, 1983. Cap. 4, p.122-155. (Grandes cientistas sociais, 39).

BUDAPESTE OPEN ACCESS INITIATIVE. **Budapest Open Access Initiative**. Budapest: BOAI, 2002. Disponível em: <http://www.opensocietyfoundations.org/openaccess/read>. Acesso em: 6 set. 2020.

BUCKLAND, M. K. Information as thing. **Journal of the American Society for Information Science - JASIS**, v. 45, n. 5, p. 351–360, 1991. Disponível em: [https://skat.ihmc.us/rid=1KR7VC4CQ-SLX5RG-5T39/BUCKLAND\(1991\)-informationasthing.pdf](https://skat.ihmc.us/rid=1KR7VC4CQ-SLX5RG-5T39/BUCKLAND(1991)-informationasthing.pdf). Acesso em: 11 maio 2020.

BURCH, S. **Sociedade da informação/ sociedade do conhecimento**. Maio 2006. Disponível em: <https://pt.scribd.com/document/470843156/Sociedade-Da-Informacao-e-Do-Conhecimento-Sally-Burch>. Acesso em: 31 maio 2023.

BYRD, J. B. *et al.* Responsible, practical genomic data sharing that accelerates research. **Nat Rev Genet.**, v. 21, n. 10, p. 615-629, 21 out. 2020. DOI: 10.1038/s41576-020-0257-5. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/32694666/#:~:text=Responsible%2C%20practical%20genomic%20data%20sharing%20that%20accelerates%20research,grapple%20with%20what%20should%20be%20required%20or%20encouraged>. Acesso em: 14 ago. 2021.

CERVI, E. U. As amostras. In: CERVI, E. U. **Manual de métodos quantitativos para iniciantes em Ciência Política**: volume 1. Curitiba: CPOP-UFPR, 2017. Cap. 5, p. 127-142.

CONFAP. **Encontre uma FAPS**. [2021]. Disponível em: <https://confap.org.br/pt/faps>. Acesso em: 15 ago. 2021.

CORRÊA, F. C. **Gestión de datos de investigación**. Barcelona: Editorial UOC, 2017. E-book. (Colección EPI Scholar).

CORRÊA, P. L. P. Modelo organizacional para gestão integrada de dados da biodiversidade brasileira. In: DIAS, G. A.; OLIVEIRA, B. M. J. F. de (org.). **Dados científicos:** perspectivas e desafios. João Pessoa: UFPB, 2019. Cap. 7, p. 153-175.

COSTA, S. Filosofia aberta, modelos de negócios e agências de fomento: elementos essenciais a uma discussão sobre o Acesso Livre à informação científica. **Ciência da Informação**, v. 35, n. 2, p. 39-50, maio/ago. 2006. Disponível em: <http://www.ibict.br/cienciadainformacao/viewarticle.php?id=926&layout=abstract>. Acesso em: 07 abr.. 2020

COSTA, M. M. **Diretrizes para uma política de gestão de dados científicos no Brasil**. Orientador: Murilo Bastos da Cunha. 2017. 288 f. Tese (Doutorado em Ciência da Informação) — Programa de Pós-Graduação em Ciência da Informação, Universidade de Brasília, Brasília, DF, 2017.

- COX, A. M.; PINFIELD, S. Research data management and libraries: current activities and future. **Sage Journals**, 28 jun. 2013. DOI: <https://doi.org/10.1177/0961000613492542>.
- CREATIVE COMMONS BR. **Cartilha CC Brasil**: o que você precisa saber sobre licenças CC. 2 fev. 2021. Disponível em: <https://br.creativecommons.net/wp-content/uploads/sites/30/2021/02/CartilhaCCBrasil.pdf>. Acesso em: 27 jun. 2023.
- CRESWELL, J. W. **Projeto de pesquisa**: métodos qualitativo, quantitativo e misto. 3. ed. Porto Alegre: Artmed, 2010.
- CRESWELL, J. W. **Investigação qualitativa e projeto de pesquisa**: escolhendo entre cinco abordagens. Tradução: Sandra Mallmann da Rosa. 3. ed. [S.l]: Penso, 2015. Disponível em: <http://integrada.minhabiblioteca.com.br/books/9788565848893>. Acesso em: 30 mar. 2020.
- CUNHA, C. M. Aaron Swartz: bandido ou herói do acesso aberto? **Revista Ibero-Americana de Ciência da Informação**, v. 13, n. 2, p. 475-479, maio 2020. DOI:10.26512/rici.v13.n2.2020.31711. Disponível em: https://www.researchgate.net/publication/352390970_Aaron_Swartz_bandido_ou_heroi_do_acesso_aberto. Acesso em: 10 fev. 2023.
- CUNHA, M. B.; COSTA, M. M. Fontes de informação sobre gestão de dados de pesquisa. **Inf. & Soc.: Est.**, João Pessoa, v. 30, n. 4, p. 1-59, out./dez. 2020.
- CUNHA, C. M.; ALMEIDA NETO, O. P. de; STACKFLETH, R. Principais métodos de avaliação psicométrica da validade de instrumentos de medida. **Rev. Atem. Saúde**, v. 14, n. 47, p. 75-83, jan./mar. 2016. DOI: <https://doi.org/10.13037/ras.vol14n47.3391>. Disponível em: https://seer.uscs.edu.br/index.php/revista_ciencias_saude/article/view/3391. Acesso em: 12 ago. 2021.
- DA COSTA, N. C. *et al.* **Lógica paraconsistente anotada**. São Paulo: Atlas, 1999.
- DALRYMPLE, D. Scientific knowledge as a global public good: contributions to innovation and the economy. In: ESANU, J. M.; UHLIR, P. F. (ed.). **The role of scientific data and information in the public domain: proceedings of a symposium**. Washington, DC: National Academy, 2003. Cap. 5, p. 35-51. Disponível em: https://books.google.com.br/books?hl=pt-BR&lr=&id=8SoXZJZ866sC&oi=fnd&pg=PR1&ots=T-5LXzoWIZ&sig=AeOBrX5ZbCTzt28qU4La760Y1JY&redir_esc=y#v=onepage&q&f=false. Acesso em: 09 mar. 2021.
- DENCKER, A. F. M. **Métodos e técnicas de pesquisa em turismo**. 4. ed. São Paulo: Futura, 2000.
- DIGITAL CURATION CENTRE. **DCC Curation Lifecycle Model**: institucional. [2018?]. Disponível em: <https://www.dcc.ac.uk/resources/curation-lifecycle-model>. Acesso em: 18 ago. 2021.
- THE ENGINEERING AND PHYSICAL SCIENCES RESEARCH COUNCIL. **EPSRC policy framework on research data**. 2018. Disponível em: <https://goo.gl/8jDswQ>. Acesso em: 05 ago. 2021.
- EUROPEAN COMMISSION. **Open innovation, open science, open world: a vision for Europe**. Luxembourg: Publications Office of the European Union, 2020. Disponível em: <https://ec.europa.eu/digital-single-market/en/news/open-innovation-open-science-open-world>

vision-europe. Access on: 10 jan. 2021.

FAPESP. **Open science Fapesp**. [2019]. Disponível em: <https://www.fapesp.br/openscience/>. Acesso em: 14 jan. 2021.

FÁVERO, L. P. *et al.* **Análise de dados**: modelagem multivariada para tomada de decisões. 5. ed. Rio de Janeiro: Elsevier, 2009.

FERNANDES L. A.; GOMES, J. M. M. Relatório de pesquisa nas Ciências Sociais: Características e modalidades de investigação. **ConTexto**, Porto Alegre, v. 3, n. 4, 2003. Disponível em: <https://seer.ufrgs.br/ConTexto/article/view/11638>. Acesso em: 14 nov. 2020.

FIX, Blair. The Legacy of Aaron Swartz: The Fight for Open Access. **Capital as power**, 19 dec. 2019. Disponível em: <https://capitalaspower.com/2019/12/the-legacy-of-aaron-swartz-the-fight-for-open-access/>. Acesso em: 07 jan. 2023.

FORCE 11. **Guiding principles for findable, accessible, interoperable and re-usable data publishing version B1.0**. [2012]. Disponível em: <https://www.force11.org/fairprinciples>. Acesso em: 01 set. 2021.

FOSTER. **Open Science Definition**. 2018a. Disponível em: <https://www.fosteropenscience.eu/foster-taxonomy/open-science-definition>. Acesso em: 15 nov. 2020.

FOSTER. Revisão por pares aberta, métricas e avaliação: capítulo 8. In: FOSTER **Introdução a Ciência Aberta**. 12 dec. 2018b. https://github.com/Open-Science-Training-Handbook/Open-Science-Training-Handbook_PT/blob/master/02IntroducaoCienciaAberta/08Revisao_por_Pares_Aberta_Metricas_Avaliacao.md. Acesso em: 15 nov. 2020.

FOUCAULT, M. O que é um autor?, **Bulletin de la Société Française de Philosophie**, ano 63, n. 3, p. 73-104, jul./set. 1969. Société Française de Philosophie, 22 de fevereiro de 1969; debate com M. de Gandillac, L. Goldmann, J. Lacan, J. d'Ormesson, J. Ullmo, J. Wahl.. Disponível em: https://edisciplinas.usp.br/pluginfile.php/276782/mod_resource/content/1/Foucault%20Michel%20-%20que%20%C3%A9%20um%20autor.pdf. Acesso em: 23 fev. 2021.

FRASER, N.; KAMER, B. **COVID-19 Preprints**. 29 jan. 2023. DOI: <https://doi.org/10.6084/m9.figshare.12033672>. Disponível em: https://github.com/nicholasmfraser/covid19_preprints. Acesso em: 22 fev. 2023.

HOOK, Daniel. Open Access surpasses subscription publication globally for the first time. **Blog Dimensions**, 24 fev. 2021. Disponível em: <https://www.dimensions.ai/blog/open-access-surpasses-subscription-publication-globally-for-the-first-time/>. Acesso em: 18 fev. 2023.

GODOY, A. S. Pesquisa qualitativa: tipos fundamentais. **Revista de Administração de Empresas**, São Paulo, v. 35, n. 3, p. 20-29, maio/jun. 1995. DOI: <https://doi.org/10.1590/S0034-75901995000300004>. Disponível em: <https://www.scielo.br/j/rae/a/ZX4cTGrqYfVhr7LvVyDBgdb/?format=pdf&lang=pt>. Acesso em: 11 ago. 2021.

GOLD, A. Cyberinfrastructure, data, and libraries, Part 1: a cyberinfrastructure primer for librarians. **D-Lib Magazine**, v. 13, n. 9/10, online, 2007. Disponível em: <http://www.dlib.org/dlib/september07/gold/09gold-pt1.html>. Acesso em: 27 ago. 2020.

GONZÁLEZ DE GÓMEZ, M. N. Para uma reflexão epistemológica acerca da Ciência da Informação. **Perspect. Cienc. Inf.**, Belo Horizonte, v. 6, n. 1, p. 5-18, jan/jun. 2001. Disponível em: <http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/433>. Acesso em: 30 ago. 2020.

GRAY, J. *et al.* Scientific data management in the coming decade. **ACM SIGMOD Record**, v. 34, n. 4, p. 34-41, 2005. Disponível em: <https://www.cs.princeton.edu/courses/archive/spr05/cos598E/bib/science-decade.pdf>. Acesso em: 14 jun. 2021.

GUIMARÃES, M. C. S. Ciência aberta e livre acesso à informação científica: tão longe, tão perto. **RECIIS – Rev. Eletron. de Comun. Inf. Inov. Saúde**, v. 8, n. 2, p.139-152, jun. 2014. Disponível em: <https://www.arca.fiocruz.br/bitstream/icict/17106/2/7.pdf>. Acesso em: 16 nov. 2019.

HALBERT, M. Prospects for research data management. In: COUNCIL ON LIBRARY AND INFORMATION RESOURCES. **Research Data Management: Principles, Practices, and Prospects**. Washington, DC: Council on Library and Information Resources, 2013. p. 1-14. Disponível em: <https://www.clir.org/wp-content/uploads/sites/6/pub160.pdf>. Acesso em: 15 ago. 2021.

HIGGINS, S. Digital curation: the emergence of a new discipline. **The International Journal of Digital Curation**, v. 6, n. 2, 2008. DOI: <https://doi.org/10.2218/ijdc.v6i2.191>. Disponível em: <http://www.ijdc.net/index.php/ijdc/article/view/184>. Acesso em: 22 ago. 2021.

HITCHCOCK, S. (Ed.). **The effect of open access and downloads ('hits') on citation impact: a bibliography of studies**. 15 set. 2004. Disponível em: <https://eprints.soton.ac.uk/354006/1/oacitation-biblio-snapshot0613.html#indexes>. Acesso em: 04 fev. 2021.

HJØRLAND, Birger. Data (with big data and database semantics). **Knowledge Organization**, v. 45, n. 8, p. 685-708, 2018. ISKO Encyclopedia of Knowledge Organization, ed. Birger Hjørland, coed. Claudio Gnoli. Disponível em: <https://www.isko.org/cyclo/data.htm>. Acesso em 09 maio. 2020.

HOURCADE, V. **O movimento da Ciência Aberta no Brasil**. Orientação: Marta Mourão Kanashiro. 2015. 144 f. Dissertação (Mestrado em Divulgação Científica e Cultural) - Instituto de Estudos da Linguagem, Universidade Estadual de Campinas, Campinas, 2015.

INSTITUTO NACIONAL DE PROPRIEDADE INDUSTRIAL. **Patentes**. Brasília, [201-?]. Disponível em: <https://www.gov.br/inpi/pt-br>. Acesso em: 20 dez. 2020.

JARDIM, J. M.; SILVA, S. C. de A.; NHARRELUGA, R. S. Análise de políticas públicas: uma abordagem em direção às políticas públicas de informação. **Perspect. Ciênc. Inf.**, v. 14, n. 1, abr. 2009. DOI: <https://doi.org/10.1590/S1413-99362009000100002>. Disponível em: <https://www.scielo.br/j/pci/a/86sqfsg3NJnCXKFmxfg5C9v/>. Acesso em: 27 jun. 2023.

JOHNSON, S. **Cultura da interface**: como o computador transforma nossa maneira de criar e comunicar. Tradução Maria Luiza X. de A. Borges. Rio de Janeiro: Jorge Zahar, 2001.

JORGE, V. de A. **Abertura e compartilhamento de dados para pesquisa nas situações de emergência em saúde pública**: o caso do vírus Zica. 2018. 266 f. Tese (Doutorado em Ciência da Informação) - Universidade Federal do Rio de Janeiro, Escola de Comunicação, Instituto Brasileiro de Informação em Ciência e Tecnologia, Rio de Janeiro, 2018. Disponível em: https://ridi.ibict.br/bitstream/123456789/993/1/Tese_VanessaJorge.pdf. Acesso em: 2 set. 2020.

- KON, F. Ciência aberta, dados abertos e código aberto. **Computação Brasil**, Especial Educação Aberta, São Paulo, n. 22, jul. 2013. Disponível em: <https://www.ime.usp.br/~kon/papers/ComputacaoBrasilKon2013.pdf>. Acesso em: 2 fev. 2021.
- LAFUENTE, A. Aprender a afectarse: un enfoque procomún del trabajo social. **Procomún y trabajo social poco a poco, paso a paso,... entre todos** - blog, 17 nov. 2014. Disponível em: <https://procomunytrabajosocial.wordpress.com/2014/11/17/aprender-a-afectarse-un-enfoque-procomunal-del-trabajo-social/>. Acesso em: 03 fev. 2021.
- LIMA, C. M. de; SANTINI, R. M. *Copyleft* e licenças criativas de uso de informação na sociedade da informação. **Ci. Inf.**, Brasília, v. 37, n. 1, p. 121-128, jan./apr. 2008. Disponível em: https://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-19652008000100011. Acesso em: 15 dez. 2020.
- LEMLE, M.; MANNHEIMER, V. Ciência aberta em história: como fazer? **Revista HCSM**, Manguinhos, ago. 2019. Disponível em: <http://www.revistahcsm.coc.fiocruz.br/ciencia-aberta-em-historia-como-fazer/>. Acesso em: 15 nov. 2019.
- LUCIAN, R.; DORNELAS, J. S. Mensuração de Atitude: Proposição de um Protocolo de Elaboração de Escalas. **RAC**, Rio de Janeiro, v. 19, 2. edição especial, art. 3, p. 157-177, ago. 2015. Disponível em: <https://www.scielo.br/pdf/rac/v19nspe2/1982-7849-rac-19-spe2-0157.pdf>. Acesso em: 15 dez. 2020.
- MACHADO, J. Dados abertos e ciência aberta. In: ALBAGLI, S.; MACIEL, M. L.; HANNUD ABDO, A. (org.). **Ciência aberta: questões abertas**. Brasília: IBICT; Rio de Janeiro: UNIRIO, 2015. p. 201-228. Disponível em: [http://livroaberto.ibict.br/bitstream/1/1060/1/Ciencia%20aberta_questoes%20abertas_PORTUGUES_DIGITAL%20\(5\).pdf](http://livroaberto.ibict.br/bitstream/1/1060/1/Ciencia%20aberta_questoes%20abertas_PORTUGUES_DIGITAL%20(5).pdf). Acesso em: 16 nov. 2020.
- MANYIKA, J., et al. **Open data: unlocking innovation and performance with liquid information**. McKinsey, 2013. Disponível em: http://www.mckinsey.com/insights/business_technology/open_data_unlocking_innovation_and_performance_with_liquid_information. Acesso em: 11 ago. 2022.
- MARTINS, G. A. Metodologias convencionais e não-convencionais e a pesquisa em administração. **Cadernos de pesquisa em administração**, São Paulo, jul.-dez. 1994. Disponível em: https://www.academia.edu/35581495/METODOLOGIAS_CONVENTIONAIS_E_N%C3%83O-CONVENTIONAIS_E_A_PESQUISA_EM_ADMINISTRA%C3%87%C3%83O. Acesso em: 21 set. 2019.
- MARTINS, H. H. T. de S. Metodologia qualitativa de pesquisa. **Educação e Pesquisa**, São Paulo, v. 30, n. 2, p. 289-300, maio/ago. 2004. Disponível em: <https://www.scielo.br/pdf/ep/v30n2/v30n2a07>. Acesso em: 08 abr. 2021.
- MARTINS, H. C. A importância da Ciência Aberta (Open Science) na pesquisa em Administração. **Rev. Adm. Contemp.**, v. 24, n. 1, jan-feb. 2020. DOI: <https://doi.org/10.1590/1982-7849rac2020190380>. Disponível em: <https://www.scielo.br/j/rac/a/xdsnKjbRg6BD6nzFXnKnVhb/>. Acesso em: 03 jun. 2023.
- MAYER-SCHONBERGER, V.; CUKIER, K. **Big data: como extrair volume, variedade, velocidade e valor da avalanche de informação cotidiana**. Rio de Janeiro: Campus : Elsevier, 2013.

- MÉNDEZ, E. “Open Science” e investigación en el siglo XXI. **Magazine**, Madrid, 13 feb. 2018. Disponível em:
<http://uc3m-magazine.uc3m.es/index.php/2018/02/13/open-science-e-investigacion-en-el-siglo-xxi/>. Acesso em: 06 nov. 2020.
- MERTON, R. K. The normative structure of science. 1942 repr. In: MERTON, R. K. **The Sociology of Science**. Chicago: University of Chicago, 1973.
- MICHEL, M. H. **Metodologia e pesquisa científica em Ciências Sociais**. 2. ed. São Paulo: Atlas, 2009.
- MOKYR, J. The commons of knowledge: a historical perspective. **The Annual Proceedings of the Wealth and Well-Being of Nations**, 2011-2012. 2008. Disponível em:
<https://www.beloit.edu/academics/economics/upton-program-forum/>. Acesso em: 16 nov. 2019.
- MONTEIRO, G.; LUCAS, E. R. de O. Dados científicos abertos: identificando o papel das políticas de gestão e das agências de fomento. **AtoZ**, Paraná, v. 8, n. 1, 2019. Disponível em:
<https://revistas.ufpr.br/atoz/article/view/67253/38725>. Acesso em: 15 nov. 2019.
- MUELLER, S. P. M. A comunicação científica e o movimento de acesso livre ao conhecimento. **Ci. Inf.**, Brasília, v. 35, n. 2, p. 27-38, maio/ago. 2006. Disponível em:
<https://www.scielo.br/j/ci/a/nGD3MkKfNxtjnnWshf3YVjP/>. Acesso em: 04 fev. 2021.
- MURRAY, J. **Hamlet no Holodeck**: o futuro da narrativa no ciberespaço. Tradução: Elissa Khouri Daher e Marcelo Cuzziol. São Paulo: Itaú Cultural, 2003.
- NIELSEN, M. **Open Science Now!** [Palestra proferida durante o TEDx Waterloo]. 2011. Disponível em:
https://www.ted.com/talks/michael_nielsen_open_science_now?language=pt-br. Acesso em: 28 abr. 2020.
- OECD. **OECD Principles and guidelines for access to research data from public funding**. Paris, 2007. Disponível em: <http://www.oecd.org/sti/inno/38500813.pdf>. Acesso em: 11 ago. 2021.
- OECD. **Promoting access to public research data for scientific, economic and social development**. Paris, 2010. Disponível em: <http://www.oecd.org/sHYPERLINK%22http://www.oecd.org/science/scienceandtechnologypolicy/38500813.pdf%22cience/scienceandtechnologypolicy/38500813.pdf>. Acesso em: 30 out. 2014.
- OECD. **Making Open Science a Reality**. Paris: OECD, 2015. (Documentos de Política da OCDE ciência, tecnologia e indústria, 25). Disponível em:
https://read.oecd-ilibrary.org/science-and-technology/making-open-science-a-reality_5jrs2f963zs1-en#page2. Acesso em: 22 abr. 2020.
- OLIVEIRA, A. C. S. de. **Desvendando a autoralidade colaborativa na e-science sob a ótica dos direitos de propriedade intelectual**. 2016. 300 f. Dissertação (Doutorado em Ciência da Informação) - Universidade Federal da Paraíba, João Pessoa, 2016. Disponível em:
<https://repositorio.ufpb.br/jspui/handle/tede/8849>. Acesso em: 04 jan. 2021.
- OUCHI, M. T. **Um estudo da Blockchain aplicado ao contexto dos dados de pesquisa**. 2019. 92 f. Dissertação (Mestre em Ciência da Informação) - Universidade Federal de São Carlos, São Paulo, 2019.

- RECKER, J. *et al.* CoreTrustSeal Trustworthy Data Repositories Requirements: Glossary 2020–2022. **Zenodo**, 20 nov. 2019. Disponível em: https://zenodo.org/record/3632563#.YS_q0PeSnIU. Acesso em: 01 set. 2021.
- PACKER, A. L.; SANTOS, S. Ciência Aberta e o novo *modus operandi* de comunicar pesquisa. **Boletim Informativo da Sociedade Brasileira de Ciência do Solo**, v. 45, n. 1, p. 17-24, 1 aug. 2019a. Disponível em: <https://www.sbcn.org.br/wp-content/uploads/2019/06/Boletim-SBCS-Volume-45-N%C3%BAmero-1.pdf>. Acesso em: 07 fev. 2021.
- PACKER, A. L.; SANTOS, S. Ciência Aberta e o novo *modus operandi* de comunicar pesquisa: parte II [online]. **SciELO em Perspectiva**, 1 aug. 2019b. Disponível em: <https://blog.scielo.org/blog/2019/08/01/ciencia-aberta-e-o-novo-modus-operandi-de-comunica-r-pesquisa-parte-ii/>. Acesso em: 08 fev. 2021.
- PARCERIA PARA GOVERNO ABERTO. **1º plano de ação nacional em Governo Aberto**. Brasília, 2013. Disponível em: <https://www.gov.br/cgu/pt-br/governo-aberto/a-ogg/planos-de-acao>. Acesso em: 21 fev. 2021.
- PARCERIA PARA GOVERNO ABERTO. **Balanço final: 2º plano de ação do Brasil**. Brasília, 2016a. Disponível em: <https://www.gov.br/cgu/pt-br/governo-aberto/central-de-conteudo/documentos/balanco-final-2-o-plano-versao-final.pdf>. Acesso em: 21 fev. 2021.
- PARCERIA PARA GOVERNO ABERTO. **3º plano de ação do Brasil**. Brasília, 2016b. Disponível em: <https://www.gov.br/cgu/pt-br/governo-aberto/central-de-conteudo/documentos/3o-plano-de-acao-versao-final.pdf>. Acesso em: 22 fev. 2021.
- PARCERIA PARA GOVERNO ABERTO. **4º plano de ação nacional em Governo Aberto**. Brasília, 2018. Disponível em: <https://www.gov.br/cgu/pt-br/governo-aberto/a-ogg/planos-de-acao/4o-plano-de-acao-brasileiro/4-plan-acao-nacional-atualizado2020.pdf>. Acesso em: 21 fev. 2021.
- PARRA, H. Z. M. Ciência cidadã: modos de participação e ativismo informacional. In: ALBAGLI, S.; MACIEL, M. L.; HANNUD ABDO, A. (org.). **Ciência aberta: questões abertas**. Brasília: IBICT; Rio de Janeiro: UNIRIO, 2015. p. 121-142. Disponível em: [http://livroaberto.ibict.br/bitstream/1/1060/1/Ciencia%20aberta_questoes%20abertas_PORTUGUES_DIGITAL%20\(5\).pdf](http://livroaberto.ibict.br/bitstream/1/1060/1/Ciencia%20aberta_questoes%20abertas_PORTUGUES_DIGITAL%20(5).pdf). Acesso em: 17 nov. 2020.
- PAVÃO, Caterina Groposo *et al.* **Acesso aberto a dados de pesquisa no Brasil**: repositórios brasileiros de dados de pesquisa: relatório 2018. Handle: 2050011959/20180801. Disponível em: <http://hdl.handle.net/2050011959/127>. Acesso em: 09 mar. 2023.
- PIOVESAN, A.; TEMPORINI, E. R. Pesquisa exploratória: procedimento metodológico para o estudo de fatores humanos no campo da saúde pública. **Rev. Saúde Pública**, v. 29, n. 4, p. 318-325, 1995. Disponível em: <http://www.scielo.br/pdf/rsp/v29n4/10>. Acesso em: 22 maio 2018.
- PIORUN, M. E. **E-Science as a Catalyst for Transformational Change in University Research Libraries**. 2013. 369 f. Tese (Doutorado em Filosofia) – Faculty of the Simmons College Graduate School of Library and Information Science. University of Massachusetts Medical School, 2013. Disponível em: <http://escholarship.umassmed.edu/cgi/viewcontent.cgi?article=1151&context=lib>. Acesso em: 10 março. 2021.

POLIT, D. F.; BECK, C. T. The content validity index: are you sure you know what's being reported? Critique and recommendations. **Research in Nursing & Health**, n. 29, p. 489–497, 2006.

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 abr. 2019.

QUIVY, R.; CAMPENHOUD, L. **Manual de Investigación en Ciencias Sociales**. México: Limusa, 2005.

REA; L. M.; PARKER, R. A. **Metodologia de pesquisa**: do planejamento à execução. São Paulo: Pioneira, 2000.

RIBEIRO, N. C.; SILVEIRA, L. da; SANTOS, S. R. de O. **Taxonomia da Ciência Aberta** [versão em português]. 29 abr. 2020. Disponível em: https://figshare.com/articles/figure/Taxonomia_da_Ci_nicia_Aberta/12124002. Acesso em: 05 jun. 2023. Originalmente publicado por FOSTER.

ROGERS, E. **Diffusion of innovations**. 3. ed. New York: Free, 1983.

SÁ, I. P. de; CAMPOS, M. L. de A. Dados científicos no Brasil no contexto da ciência aberta: uma análise preliminar. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 18., 2017, Marília. **Anais...** Marília: Unesp, 2017. Disponível em: http://enancib.marilia.unesp.br/index.php/XVIII_ENANCIB/ENANCIB/paper/viewFile/281/180. Acesso em: 05 jun. 2020.

SALES, L. F. **Integração semântica de publicações científicas e dados de pesquisa: proposta de modelo de publicação ampliada para a área de ciências nucleares**.

Orientadores: Rosali Fernandez de Souza e Luís Fernando Sayão. 2014. 264 f. Tese (Doutorado em Ciência da Informação) – Programa de Pós-Graduação em Ciência da Informação, Instituto Brasileiro de Informação e Tecnologia, Universidade Federal do Rio de Janeiro, 2014.

SANCHES, C.; MEIRELES, M.; SORDI, J. de. O. Análise qualitativa por meio da lógica paraconsistente: método de interpretação e síntese de informação obtida por escalas Likert. In: ENCONTRO DE ENSINO E PESQUISA EM ADMINISTRAÇÃO E CONTABILIDADE, 3., 2011, João Pessoa. **Anais eletrônicos...** João Pessoa: Anpad, 2011. Disponível em: http://www.anpad.org.br/diversos/trabalhos/EnEPQ/enepq_2011/ENEPEQ221.pdf. Acesso em: 30 set. 2021.

SAYÃO, L. F; SALES, L. F. Curadoria digital: um novo patamar para a preservação de dados digitais de pesquisa. **Inf. & Soc.: Estudos**, João Pessoa, v. 22, n. 3. p.179-191, set./dez. 2012. Disponível em: <https://www.proquest.com/docview/1493879494?pq-origsite=gscholar>. Acesso em: 18 ago. 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/images/CIN/PDFs/GUIA_DE_DADOS_DE_PESQUISA.pdf. Acesso em: 18 ago. 2021.

SAYÃO, L. F.; SALES, L. F. Algumas considerações sobre os repositórios digitais de dados de pesquisa. **Informação & Informação**, v. 21, n. 2, p. 90–115, 2016. Disponível em: <http://www.uel.br/revistas/uel/index.php/informacao/article/view/27939>. Acesso em: 20 set. 2019.

- SAYÃO, L. F.; SALES, L. F. Afinal, o que é dado de pesquisa. **Biblos**, v. 34, n. 2, p. 32-51, jul./dez. 2020. DOI: <https://doi.org/10.14295/biblos.v34i2.11875>. Disponível em: <https://periodicos.furg.br/biblos/article/download/11875/8426/38>. Acesso em: 17 ago. 2021.
- SCHWARTZMAN, S. Os mitos da ciência. In: SCHWARTZMAN, S. **Ciência, universidade e ideologia**: a política do conhecimento [online]. Rio de Janeiro: Centro Edelstein, 2008. Cap. 2, p. 18-38. Disponível em: <http://books.scielo.org/id/mny2p/pdf/schwartzman-9788599662502-03.pdf>. Acesso em: 16 nov. 2019.
- SEMERER, A. R.; PINTO, A. L. Os diferentes conceitos de dados de pesquisa na abordagem da biblioteconomia de dados. **Ci. Inf.**, Brasília, DF, v. 48, n. 1, p.130-129, jan./abr. 2019. Disponível em: <http://revista.ibict.br/ciinf/article/view/4461>. Acesso em: 14 ago. 2021.
- SETZER, V. W. Dado, informação, conhecimento e competência. **DataGramZero - Revista de Ciência da Informação**, dez. 1999. Disponível em: <https://www.ime.usp.br/~vwsetzer/datagrama.html>. Acesso em: 15 maio 2021.
- SILVA, F. C. C da. **Gestão de dados científicos**. Rio de Janeiro: Interciência, 2019.
- SILVA, P. F. P da. **As políticas de Open Data em Portugal**: análise da sua implementação e impacto. 2017. 153 f. Dissertação (Mestrado em Ciência da Informação) – Departamento de Filosofia, Comunicação e Informação da Faculdade de Letras da Universidade de Coimbra, 2017. DOI 10.13140/RG.2.2.19553.20326. Disponível em: https://www.researchgate.net/publication/322011884_As_Politicas_de_Open_Data_em_Portugal_analise_da_sua_implementacao_e_impacto/figures?lo=1. Acesso em: 19 jan. 2023.
- SILVA, F. C. C. da; SILVEIRA, L. da. O ecossistema da Ciência Aberta. **Transinformação**, v. 31, e190001, 2019. Disponível em: <http://dx.doi.org/10.1590/2318-0889201931e190001>. Acesso em: 27 nov. 2020.
- SOCIENTIZE CONSORTIUM. **Green Paper on Citizen Science**. [S.l.]: European Commission, 2013. Disponível em: <http://ec.europa.eu/digital-agenda/en/news/green-paper-citizen-science-europe-towards-society-empowered-citizens-and-enhanced-research-0>. Acesso em: 1 out. 2020.
- SOLLA PRICE, Derek de. **Little science, big science**. New York: Columbia University, 1969.
- SWARTZ, A. **Guerilla Open Access Manifesto**. Erema, Italy, july 2008. Disponível em: <https://ia600605.us.archive.org/15/items/GuerillaOpenAccessManifesto/Goamjuly2008.pdf>. Acesso em: 14 nov. 2022.
- TENNANT, J. et al. The academic, economic and societal impacts of Open Access: an evidence-based review. **F1000Research**, v. 3, n. 5, p. 1-55, 10 abr. 2016. Disponível em: <https://f1000research.com/articles/5-632/v3>. Acesso em: 12 fev. 2021.
- TENOPIR, C. et al. Research data services in European Academic Research Libraries. **LIBER Quarterly**, v. 27, n.1, p. 23–44, 2017. Disponível em: <https://www.liberquarterly.eu/articles/10.18352/lq.10180/>. Acesso em: 02 jun. 2021.
- TORINO, E.; ROA-MARTÍNEZ, S. M.; VIDOTTI, S. A. B. G. Dados de pesquisa: disponibilização ou publicação?. In: SHINTAKU, M.; SALES, L. F; COSTA, M. (org). **Tópicos sobre dados abertos para editores científicos**. Botucatu, SP: ABEC, 2020. p. 183-201. DOI: 10.21452/ 978-85-93910-04-3.cap15. Disponível em: <http://repositorio.utfpr.edu.br/jspui/bitstream/1/4725/1/dadospesquisapublicacao.pdf>. Acesso

em: 22 jun. 2022.

UFBA. Plano de dados abertos da Ufba está disponível para consulta pública. 10 dez. 2018. Disponível em:

https://ufba.br/ufba_em_pauta/plano-de-dados-abertos-da-ufba-est%C3%A1-dispon%C3%A1vel-para-consulta-p%C3%BAblica#:~:text=O%20Plano%20de%20Dados%20Abertos%20%C3%A9%20um%20instrumento,%C3%A0%20informa%C3%A7%C3%A3o%20e%20do%20est%C3%ADmulo%20%C3%A0%20p. Acesso em 14 out. 2022.

UFBA. Plano de desenvolvimento institucional 2018-2022. Salvador, 2017. Disponível em: <https://proplan.ufba.br/sites/proplan.ufba.br/files/pdi-2018-2022.pdf>. Acesso em: 10 mar. 2023.

UK RESEARCH COUNCILS. Common principles on data policy. 2011. Disponível em: <http://www.rcuk.ac.uk/research/Pages/DataPolicy.aspx>. Acesso em: 31 jan. 2014.

UK MULTI-STAKEHOLDER GROUP. Concordat on open research data. 2016.

Disponível em:

<https://www.ukri.org/wp-content/uploads/2020/10/UKRI-020920-ConcordatonOpenResearchData.pdf>. Acesso em: 05 ago. 2021.

UNESCO. Diretrizes para recursos educacionais abertos (REA) no ensino superior.

2015. Disponível em:

http://www.unesco.org/new/fileadmin/MULTIMEDIA/HQ/CI/CI/pdf/publications/oer_guidelines_pt.pdf. Acesso em: 12 jan. 2021.

UNESCO. Ciência, tecnologia e inovação no Brasil. Brasília, [2020?]. Disponível em: Ciência, tecnologia e inovação no Brasil (unesco.org). Acesso em: 14 jan. 2021.

UZWYSHYN, R. Research Data Repositories: the what, when, why, and How. **Computers in Libraries**, v. 36, n. 3, abr. 2016. Disponível em:

<https://www.infotoday.com/cilmag/apr16/Uzwyshyn--Research-Data-Repositories.shtml>. Acesso em: 11 ago. 2021.

VANZ, Samile Andrea de Souza et al. **Acesso aberto a dados de pesquisa no Brasil: práticas e percepções dos pesquisadores: relatório 2018.** Disponível em: <http://hdl.handle.net/10183/185195>. Acesso em: 09 mar. 2023.

VEIGA, V. S. de O. **Percepção dos pesquisadores portugueses e brasileiros da área de Neurociências quanto ao compartilhamento de artigos científicos e dados de pesquisa no acesso aberto verde: custos, benefícios e fatores contextuais.** 2017. 294 f. Tese (Doutorado em Informação e Comunicação em Saúde) - Instituto de Comunicação e Informação Científica e Tecnológica em Saúde, Fundação Oswaldo Cruz, Rio de Janeiro, 2017.

VEIGA, V. S. de O. et al. Compartilhamento de dados de pesquisa na Fiocruz: diagnóstico e percepção do pesquisador. **Ci.Inf.**, Brasília, DF, v. 48 n. 3 (Supl.), p. 319-327, set./dez. 2019. Disponível em: <http://revista.ibict.br/ciinf/article/view/4999/4464>. Acesso em: 14 jul. 2021.

VERGARA, S. C. **Projetos e relatórios de pesquisa em Administração.** São Paulo: Atlas, 2003.

WANG, X. et al. The open access advantage considering citation, article usage and social media attention. **Scientometrics**, v. 103, n. 2, p. 555-564, 2015. Disponível em: <link.springer.com/article/10.1007%2Fs11192-015-1547-0>. Acesso em: 04 jan. 2021.

WOELFLE, M.; OLLIARO, P.; TODD, M. H. Open science is a research accelerator. **Nature Chemistry**, v. 3, n. 10, p. 745–748, 23 set. 2011. DOI:10.1038/nchem.1149. Disponível em: <https://www.nature.com/articles/nchem.1149>. Acesso em: 29 out. 2019.

WORLD WEB CONSORTIUM BRASIL. **Manual dos dados abertos**: governo. 2011. Disponível em: http://www.w3c.br/pub/Materiais/PublicacoesW3C/Manual_Dados_Abertos_WEB.pdf. Acesso em: 28 jan. 2022.

YIN, R. K. **Estudo de caso**: planejamento e métodos. 5. ed. Porto Alegre: Bookman, 2015.

YUSOFF, M. S. B. ABC of content validation and content validity index calculation. **Resource**, v. 11, n. 2, p. 49-54, 2019. Disponível em: https://www.researchgate.net/publication/334134963_ABC_of_Content_Validation_and_Content_Validity_Index_Calculation. Acesso em: 29 ago. 2021.

APÊNDICE A - Apresentação e Termo de Consentimento Livre e Esclarecido

Prezado(a) professor (a), você está sendo convidado(a) a participar voluntariamente de uma pesquisa científica cujo objetivo é: conhecer a percepção dos professores pesquisadores no que tange à gestão de dados e às ações desenvolvidas pela minha Instituição para a implantação de uma política de gestão de dados de pesquisas no contexto da Ciência Aberta. Tendo por orientador o Prof. Dr. Raymundo das Neves Machado (raymacha@ufba.br) e discente, Normaci Correia dos Santos Sena (normaci@ufba.br).

A Lei de Acesso à Informação vigora no Brasil desde 2012 e marca o início da obrigatoriedade do acesso à informação pública. Resultante dessa lei, o Plano de Dados Abertos (PDA) prevê a abertura de dados; isto é, a disponibilização de dados em formato bruto, os quais as pessoas possam livremente acessar, manipular e reutilizar, ampliando, com isso, a participação social. Consequentemente, o movimento pela ciência aberta se insere nesse quadro de tensão entre, por um lado, novas formas de produção colaborativa, interativa e compartilhada da informação, do conhecimento e da cultura. E, por outro, mecanismos de captura e privatização desse conhecimento que é coletiva e socialmente produzido.

Assim, em um contexto de pesquisa e inovação e considerando que as universidades, enquanto instituições de pesquisa, precisam adaptar-se, preparar-se e atender a essa nova demanda informacional, e que os cientistas também têm como função publicizar as suas atividades de pesquisa, uma das características desse movimento de ação colaborativa de acesso livre é a definição de políticas, incluindo a gestão de dados e tecnologias digitais, tendo como centro as pessoas, é mister conhecer a opinião dos professores pesquisadores a respeito dessa temática e os principais desafios.

Agradecemos a sua colaboração ao responder às questões apresentadas. Gostaríamos de informar que esta pesquisa é apenas para fim acadêmico e, desde já, asseguramos a preservação dos dados coletados.

Tempo estimado de Resposta: 8 a 10 min.

- () Estou de acordo
- () Não estou de acordo

APÊNDICE B - Questionário aplicado aos professores de pós-graduação

Este questionário é um instrumento para pesquisa de campo, que visa conhecer a percepção dos professores pesquisadores no que tange à gestão de dados e às ações da UFBA desenvolvidas para a implantação de uma política de gestão de dados de pesquisas no contexto da Ciência Aberta.

Forneça a sua opinião franca sobre as perguntas abaixo relacionadas

1 – Dados pessoais - Idade:

- 18 a 28
- 29 a 39
- 40 ou 50
- 51 ou mais

2 - Área de conhecimento

- Ciências Humanas
- Ciências Exatas
- Ciências da Saúde/Biológicas
- Letras
- Artes

3 - Tempo de trabalho na instituição.

- Menos de 10 anos
- 11 a 25
- 26 ou 36
- 37 ou mais
- Outros

As perguntas a seguir referem-se à sua opinião a respeito de gestão de dados científicos e Ciência Aberta.

Marque apenas uma alternativa, de acordo com o seu grau de concordância ou discordância para cada proposição. Por exemplo, ao sinalizar o "Discordo Totalmente" demonstra que a proposição não contempla de maneira alguma o critério estabelecido; já o "Concordo Totalmente", a proposição satisfaz completamente o critério.

Categoria	Proposição	Discordo totalmente	Discordo parcialmente	Indiferente	Concordo parcialmente	Concordo totalmente
Ciência Aberta	1 - Compreendo a relevância da Ciência Aberta para o avanço da ciência.	[]	[]	[]	[]	[]
	2 - Devido ao grande volume de dados produzidos na minha pesquisa, utilizo ferramentas computacionais denominadas sistemas de gerência de workflows científicos (por exemplo, o Kepler e o Taverna) a fim de facilitar ou automatizar a gestão dos meus dados e, principalmente, sustentar todo o ciclo de pesquisa.					
Gestão de Dados de pesquisa	3 - Preciso de fonte de dados confiáveis para o projeto de pesquisa.	[]	[]	[]	[]	[]
	4 - A minha Instituição possui infraestrutura para a gestão dos meus dados.	[]	[]	[]	[]	[]
	5 - Faltam diretrizes para a reutilização de dados na minha Instituição.	[]	[]	[]	[]	[]
	6 - Percebo dificuldades para implementação da política de gestão de dados científicos na minha Instituição.	[]	[]	[]	[]	[]
Atitude e valores em relação a abertura dos dados	7 - Acredito que aumenta a credibilidade da minha pesquisa.	[]	[]	[]	[]	[]
	8 - Contribui para o fortalecimento do	[]	[]	[]	[]	[]

	avanço científico já que esses dados podem ser utilizados por outros pesquisadores em suas pesquisas.					
	9 - Contribui com a transparência na conduta da pesquisa.	[]	[]	[]	[]	[]
	10 - A abertura de dados de pesquisa contribui para uma maior integridade na pesquisa.	[]	[]	[]	[]	[]
	11 - Contribui para uma reproduzibilidade da pesquisa.	[]	[]	[]	[]	[]
Motivação para a abertura	12 - Amplia a acessibilidade da minha pesquisa, garantindo que o material seja acessível de maneira estável por longo tempo.	[]	[]	[]	[]	[]
	13 - Possibilita maior confiabilidade já que o material em repositório tem credibilidade.	[]	[]	[]	[]	[]
	14 - Aumenta o prestígio em minha área e, consequentemente, reconhecimento profissional.	[]	[]	[]	[]	[]
	15 - Amplia a visibilidade da minha pesquisa.	[]	[]	[]	[]	[]
	16 - Sou obrigado(a) a abrir por causa da minha agência financiadora.	[]	[]	[]	[]	[]
Necessidades dos	17 - Apoio técnico na elaboração do plano de gestão de	[]	[]	[]	[]	[]

pesquisadores	dados, ou seja, identificação de normas e esquema disciplinares de descrição de dados.					
	18 - Necessito de serviço de armazenamento e preservação, do tipo software que possa contribuir no armazenamento e preservação dos meus dados.	[]	[]	[]	[]	[]
	19 - Necessito de backup de arquivos.	[]	[]	[]	[]	[]
	20 - Preciso de formação e/ou consultoria na gestão de dados de pesquisa.	[]	[]	[]	[]	[]
Agências de Fomento	21 - A agência de fomento exige um plano de gestão de dados no projeto de sua pesquisa, mas não sei elaborar esse documento.					
	22 - Compartilho os dados para atender a demanda da agência de fomento.	[]	[]	[]	[]	[]
Direito autoral/ Proteção de dados	23 - Meus dados são sensíveis (potencialmente sigilosos).	[]	[]	[]	[]	[]
	24 - Meus dados são ostensivos (sem sigilo).	[]	[]	[]	[]	[]
	25 - Meus dados podem ou geram patentes.	[]	[]	[]	[]	[]
	26 - Meus dados são pessoais.	[]	[]	[]	[]	[]
	27 - Meus dados têm parceria interinstitucionais.	[]	[]	[]	[]	[]

Profissional capacitado e/ou Biblioteca	28 - Sinto necessidade de profissional capacitado para orientar em todo processo da gestão dos meus dados científicos, desde a gênese até o armazenamento e preservação.	<input type="checkbox"/>				
	29 - Percebo que a minha Instituição tem contribuído para a formação de profissional capaz de realizar o gerenciamento de dados científicos.	<input type="checkbox"/>				
	30 - A Biblioteca poderia gerenciar os meus dados científicos por meio de um repositório de dados.	<input type="checkbox"/>				
	31 - Não é papel da Biblioteca gerenciar os meus dados e sim eu, pesquisador. Cabe uma Assessoria fornecida pela Pró-reitoria.	<input type="checkbox"/>				
Razões para não compartilhar dados de pesquisa	32 - Não sei como compartilhar dados. Preciso de suporte assessoria).	<input type="checkbox"/>				
	33 - Meus dados não podem estar abertos, pois contêm informação sigilosa.	<input type="checkbox"/>				
		<input type="checkbox"/>				
	34 - Mau uso dos dados pode causar uma má interpretação por outros pesquisadores.	<input type="checkbox"/>				
	35 - Se eu compartilhar, meus dados podem ser usurpados por outros	<input type="checkbox"/>				

	pesquisadores.					
Repositório de dados	36 - Não confio na transparência e na autenticidade dos dados brutos em repositório de dados.	[]	[]	[]	[]	[]
	37 - Não tenho motivos/vantagens que incentivam a publicação de dados em um repositório de acesso aberto.	[]	[]	[]	[]	[]
	38 - Necessito de serviços de compartilhamento e acesso a dados.	[]	[]	[]	[]	[]
	39 - Não tenho conhecimento de nenhum repositório para fazer o compartilhamento.	[]	[]	[]	[]	[]
	40 - Não compartilho porque meus dados têm restrição contratual.	[]	[]	[]	[]	[]
	41 - Não compartilho porque posso perder uma boa publicação, caso eu compartilhe os dados	[]	[]	[]	[]	[]

Observações finais!

O espaço abaixo é reservado para que você possa fazer sugestões, críticas ou informações complementares sobre esse instrumento de pesquisa.

Muito obrigada pela participação!

APÊNDICE C: Projeto piloto de Plano de Gestão de Dados de Pesquisa para a UFBA

GUIA INTRODUTÓRIA

A proposta da Política de Gestão de Dados de Pesquisa (PGD) da Universidade Federal da Bahia é resultante do trabalho de pesquisa de Normaci Correia dos Santos Sena, sob orientação do Prof. Dr. Raymundo das Neves Machado, no Programa de Pós-Graduação em Ciência da Informação da Universidade Federal da Bahia (UFBA).

Este é um documento norteador e abrangente para a Universidade Federal da Bahia com objetivo de orientar a aplicação de procedimentos necessários para a pesquisa aberta e a gestão dos dados de pesquisa científica na UFBA. Descreve a respeito das responsabilidades da universidade e da comunidade acadêmica sobre gestão e preservação dos dados de pesquisa científica oriundos no cerne da Instituição, quer sejam atuais ou futuros. As diretrizes aqui expostas aplicam-se a todas as pesquisas realizadas por docentes, discentes e técnico-administrativos em educação da UFBA.

Diante o cenário científico internacional e nacional, é imprescindível a adesão da universidade ao movimento da Ciência Aberta, pois este movimento promove a união entre pesquisadores e a sociedade civil, fortalecendo o papel social das universidades. É estratégico definir os padrões administrativos e a forma de governança dos dados de pesquisa científica (todo o ciclo de vida dos dados) a fim de assegurar que os dados recebam o tratamento adequado para ser reproduzidos, compartilhados, acessados e reutilizados por outros pesquisadores e pelo público em geral, quando forem disponibilizados abertamente.

Os dados de pesquisa são produtos de pesquisa que cobrem uma ampla gama de tipos de registros e podem ser estruturados e armazenados em vários formatos de arquivos, ou seja, compreendem todos os tipos de objetos produzidos em qualquer fase do ciclo de vida da pesquisa. É todo dado produzido no decurso da pesquisa que geram dados provenientes da pesquisa e que pode estar em qualquer suporte (questionário, entrevista, áudios, vídeos, cadernos de laboratório etc). Esses dados precisam ser disponibilizados em Repositórios de Dados a fim de garantir o acesso aberto aos de pesquisa e contribuir com a Ciência Aberta uma vez que propõe a prática da ciência realizada de modo aberto, colaborativo e transparente.

Sendo assim, é relevante que todo projeto de pesquisa construído na universidade realize de maneira eficaz a gestão dos dados de pesquisa (*dataset*), sendo indicado que o pesquisador elabore o Plano de Gestão de Dados, ou Data Management Plan (DMP) – documento formal que descreve os dados produzidos no decurso de um projeto de pesquisa e pormenoriza estratégias de gerenciamento de dados que são implementados durante e após a fase de desenvolvimento da pesquisa. Além disso, o PGD, detalha os aspectos de

gerenciamento quando os dados estiverem sendo coletados, organizados, documentados e compartilhados e preservados.

Sobre os direitos de exclusividade na reutilização ou publicação dos dados de pesquisa científica desenvolvidas no domínio da UFBA, sugere-se que os dados tenham recebido o licenciamento adequado e que seja disponibilizado abertamente para o reuso, antes de ser entregues a agentes comerciais, a não ser em situação que tenha condição imposta pelo financiador da pesquisa ou alguma questão ética, legal ou contratual.

A proposta objetiva criar políticas de gestão de dados de pesquisa a fim de nortear os pesquisadores e a comunidade acadêmica sobre a forma adequada de gerenciar os seus dados. Essa proposta está em consonância com o Plano de Desenvolvimento da UFBA (PDI 2018-2022) e tem por base as inúmeras leituras realizadas para compor o referencial teórico deste estudo, ou seja, as referências nacionais e internacionais que recomendam a criação de políticas norteadoras para a implantação do plano de gestão de dados de pesquisa nas universidades em parceria com as agências de fomento, pesquisadores e profissionais da informação, como por exemplo, FOSTER, a Universidade de São Paulo, entre outros; bem como a avaliação de ferramentas automatizadas utilizadas para a criação de Plano de Gestão de Dados que já disponibiliza modelos de PGD de universidades e agência de fomentos, tais como DPMTool⁶². Como referência também, utilizou-se do projeto e das recomendações do DMP ICPR⁶³, e da Base de Dados Científicos da UFPR⁶⁴ e da Universidade de São Paulo⁶⁵. Cabe lembrar que as etapas do plano se baseiam nos Princípios FAIR (*Findable, Accessible, Interoperable, Reusable*), pois estes definem princípios para que os dados científicos sejam localizáveis, acessíveis, interoperáveis e reusáveis.

JUSTIFICATIVA

O estudo realizado na UFBA mostrou as necessidades dos pesquisadores no desenvolvimento de ações que possam norteá-los para um gerenciamento adequado dos dados de pesquisa. Percebendo a inexistência de políticas com essa finalidade, permitiu a construção dessa proposta a fim de ajudar os pesquisadores a adquirir competências necessárias para a gestão de seus dados. Isso se dará por meio de cursos, debates e políticas desenvolvidas na Instituição. Todas essas ações possibilitarão o pesquisador a se sentir capaz de compreender o ciclo de vida de dados de pesquisa e a construir o seu plano de dados de pesquisa. Para tanto, se faz necessário a parceria entre a pró-reitoria de pesquisa, setor de informática e biblioteca. A fim de criar um repositório e auxiliar na construção dos metadados e plano de gestão de

⁶² Disponível em: (<https://dmponline.dcc.ac.uk>)

⁶³ Disponível em: <https://goo.gl/TyrNLZ>

⁶⁴ Disponível em: <https://bdc.c3sl.ufpr.br/>

⁶⁵ Gestão de dados, disponível em: fapesp.br)

dados.

OBJETIVOS

- Desenvolver políticas e diretrizes para a Gestão de Dados de Pesquisa na UFBA;
- Capacitar, treinar e preparar equipe bibliotecário para auxiliar o pesquisador no processo necessário à execução de planos de gestão de dados;
- Firmar parcerias com outras universidades, agência de fomento entre outros a fim de fortalecer e contribuir com o acesso aberto na Instituição que coadunem com a gestão de dados de pesquisa e suas vertentes.

Responsabilidades e atribuições⁶⁶

No que se refere às atribuições e responsabilidades relativos á gestão de dados de pesquisa, sugere-se:

Biblioteca:

- Desenvolver serviços de dados e apoio ao pesquisador;
- Prestar consultoria, fornecer aconselhamento, orientação e ações de educação, treinamento, informação e congêneres à comunidade acadêmica da UFBA no que concerne à gestão de dados de pesquisa e aspectos relacionados, a exemplo da Ciência Aberta;
- Formar, capacitar, treinar e preparar a equipe de trabalho das bibliotecas que prestarão os serviços de dados, curadoria e de apoio ao pesquisador na UFBA;
- Intermediar soluções de armazenamento seguro para dados de pesquisa dinâmicos e estáticos em cooperação com a Superintendência de Tecnologia da Informação (STI);
- Firmar parcerias e redes de trabalho na própria instituição com outros setores e os cursos, departamentos, Pró-Reitoria de Pesquisa e Pós-Graduação e Pró-Reitoria de Graduação para promover e fortalecer o gerenciamento dos dados de pesquisa na UFBA;
- Formar e coordenar uma rede de administradores de dados (bibliotecários de cada faculdade da universidade);
- Realizar estudos com vistas a diagnosticar, mapear e sugerir melhorias para fornecer uma infraestrutura de dados na universidade.

⁶⁶ Essas informações foram baseadas na pesquisa de Juliana Lima, da Universidade Federal do Ceará, constante no referencial teórico e nas referências.

Superintendência de Tecnologia da Informação (STI):

- Fornecer infraestrutura adequada, padronizada, certificada, robusta e de alta qualidade para facilitar o processo de gerenciamento dos dados de pesquisa;
- Fornecer espaço de armazenamento exclusivamente para os dados de pesquisas desenvolvidas na UFBA;
- Fornecer o gerenciamento, controle de acesso aos dados e mecanismos de segurança da informação em conformidade com as diretrizes legais da área e do país.

Programas de Pós-Graduação *Stricto Sensu* (Mestrado e Doutorado) e Programas de Graduação:

- Solicitar a colaboração dos orientadores para supervisionar os seus alunos do mestrado e doutorado e assegurar o cumprimento da devida gestão dos dados de suas pesquisas desde o início ao final do projeto, principalmente com o preenchimento do Plano de Gestão de Dados (PGD);
- Informar ao orientador e aluno de que devem garantir, que os dados de pesquisa serão repassados para a universidade antes de sua saída da instituição por motivo de conclusão do curso;
- Informar que o aluno e o orientador são responsáveis por assegurar que os dados de terceiros, não serão repassados aos financiadores de pesquisa, agências de fomento ou qualquer outro sem a expressa autorização;
- Incentivar o seguimento das diretrizes e orientações fornecidas pelo serviço de dados e apoio ao pesquisador desempenhado pela biblioteca;
- Sugerir a participação do corpo docente e discente em pelo menos uma capacitação e/ou treinamento presencial ou a distância sobre o gerenciamento dos dados de pesquisa oferecidos pela biblioteca.

Docentes/Orientadores (Graduação, Mestrado e Doutorado):

- Supervisionar os seus alunos do mestrado e doutorado e assegurar o cumprimento da devida gestão dos dados de suas pesquisas desde o início ao final do projeto, principalmente com o preenchimento do Plano de Gestão de Dados (PGD) e que este deve ser atualizado regularmente;
- O orientador e aluno devem garantir que os dados de pesquisa serão repassados para a universidade antes de sua saída da instituição por motivo de conclusão do curso;

- O aluno e o orientador são responsáveis por assegurar que os dados de terceiros não serão repassados aos financiadores de pesquisa, agências de fomento ou qualquer outro sem a expressa autorização;
- Seguir e fazer com que sejam seguidas as diretrizes e orientações fornecidas pelo serviço de dados e apoio ao pesquisador desempenhado pela biblioteca;
- Participar de pelo menos uma capacitação e/ou treinamento presencial ou a distância sobre o gerenciamento dos dados de pesquisa oferecidos pela biblioteca;
- Garantir que os alunos participem, em pelo menos uma capacitação e/ou treinamento presencial ou a distância sobre o gerenciamento dos dados de pesquisa oferecidos pela biblioteca;
- Solicitar, a qualquer tempo, cursos, capacitações, treinamentos, consultoria e aconselhamento junto à biblioteca mediante agendamento;
- Elaborar um Plano de Gestão de Dados para cada pesquisa ou projeto desenvolvido, pelo menos nos primeiros 12 meses de seu início e procedendo à sua atualização ao longo do processo até o seu término;
- Garantir que os dados de pesquisa, códigos e quaisquer outros materiais necessários para os resultados de pesquisa sejam reproduzíveis (quando for o caso), que sigam os princípios FAIR (dados fáceis de encontrar, acessíveis, interoperáveis e reutilizáveis), documentados e compartilhados em um repositório de dados de livre escolha do pesquisador ao final do projeto, disponíveis por pelo menos 10 anos. Nos casos em que o compartilhamento não for possível, a justificativa deve constar expressamente no Plano de Gestão de Dados e ser informada no manuscrito final do trabalho;
- Caso os dados não sejam disponibilizados em um repositório, deve ser assegurada a garantia de publicação dos metadados relevantes em um repositório de dados e em quaisquer publicações de pesquisa resultantes do projeto em que conste uma declaração expressa em que conste a explicação sobre o material existente e seus conjuntos de dados adicionais, inclusive deve informar porque o acesso é restrito, fornecer contato para possíveis pedidos de informação, informar quem poderá utilizar e em que circunstâncias.

Discentes (Graduação, Mestrado e Doutorado):

- Assegurar o cumprimento da devida gestão dos dados de suas pesquisas desde o início ao final do projeto, principalmente com o preenchimento do Plano de Gestão de Dados (PGD) e que este deve ser atualizado regularmente;

- Deve garantir o repasse e entrega dos dados de pesquisa para a universidade antes de sua saída da instituição por motivo de conclusão do curso;
- Responsabilizar-se e assegurar que os dados de terceiros não serão repassados aos financiadores de pesquisa, agências de fomento ou qualquer outro sem a expressa autorização;
- Elaborar um Plano de Gestão de Dados para cada pesquisa ou projeto desenvolvido, pelo menos nos primeiros 12 meses de seu início e procedendo à sua atualização ao longo do processo até o seu término;
- Garantir que os dados de pesquisa, códigos e quaisquer outros materiais necessários para os resultados de pesquisa sejam reproduzíveis (quando for o caso), que sigam os princípios FAIR (dados fáceis de encontrar, acessíveis, interoperáveis e reutilizáveis), documentados e compartilhados em um repositório de dados à livre escolha do pesquisador ao final do projeto, disponíveis por pelo menos 10 anos. Nos casos em que o compartilhamento não for possível, a justificativa deve constar expressamente no Plano de Gestão de Dados e ser informada no manuscrito final do trabalho;
- Caso os dados não sejam disponibilizados em um repositório, deve ser assegurada a garantia de publicação dos metadados relevantes em um repositório de dados e em quaisquer publicações de pesquisa resultantes do projeto em que conste uma declaração expressa em que conste a explicação sobre o material existente e seus conjuntos de dados adicionais, inclusive deve informar porque o acesso é restrito, além de fornecer contato para possíveis pedidos de informação, informar quem poderá utilizar e em que circunstâncias;
- Seguir as diretrizes e orientações fornecidas pelo serviço de dados e apoio ao pesquisador desempenhado pela biblioteca;
- Participar de pelo menos uma capacitação e/ou treinamento presencial ou a distância sobre o gerenciamento dos dados de pesquisa oferecidos pela biblioteca;
- Solicitar, a qualquer tempo, cursos, capacitações, treinamentos, consultoria e aconselhamento junto à biblioteca e mediante agendamento.

Responsabilidades gerais dos pesquisadores (docentes, técnico-administrativos e discentes):

- O pesquisador irá escolher quais dados deseja compartilhar;

- O pesquisador pode escolher o repositório de sua preferência para o depósito dos dados;
- Não devem compartilhar dados sensíveis, registros de pacientes de qualquer natureza, ou todo o tipo de dado que possa causar algum dano a outrem, ou ainda que seja objeto de patente e afins;
- Os pesquisadores devem respeitar as leis e aspectos éticos relacionados às pesquisas, aos seres humanos, objetos ou animais;
- Os pesquisadores devem providenciar e garantir os meios necessários para que o armazenamento dos dados seja feito de forma segura;
- Compete ao pesquisador assegurar-se de que estão sendo aplicados e cumpridos os termos de utilização e licenciamento para o compartilhamento dos dados;
- Os interesses dos participantes de pesquisa devem ser assegurados, considerando as normas éticas da pesquisa, do Comitê de Ética em Pesquisa e a legislação vigente;
- Compete aos pesquisadores publicar seus dados de pesquisa em formatos digitais apropriados (não-proprietários) sempre que possível a fim de facilitar o reuso dos dados;
- Caso o pesquisador se desligue da universidade durante o período de realização da pesquisa ou após a sua conclusão, deverá transmitir a responsabilidade da gestão dos dados antes de sua saída expressa e documentada no Plano de Gestão de Dados. A referida transferência pode implicar em um acordo entre as partes para transferir essa responsabilidade para a outra instituição que o pesquisador está se filiando.

Normativos e legislações vigentes

- Lei de Acesso à Informação (Lei nº 12.527);
- Lei Geral de Proteção de Dados (Lei nº 13.709/2018);
- Plano de Desenvolvimento da UFBA (PDI 2018-2022);
- Política de Dados Aberto no Executivo Federal (Decreto Federal nº 8.777/ 2016);
- Plano de Dados Abertos da UFBA (PDA 2019-2020).

Ferramentas para a criação do plano de dados

- DPMTool (<https://dmptool.org>);
- DCC (<https://dmponline.dcc.ac.uk>).

Normas para a Gestão de Dados de Pesquisas Científicas na Universidade Federal da Bahia.

A Universidade Federal da Bahia disponibilizará aos seus docentes um repositório institucional para armazenamento de dados científicos, denominado Repositório de Dados Científicos da UFBA.

São elegíveis para uso do Repositório de Dados Científicos da UFBA os docentes, discentes e outros que tenham vínculo com a Universidade Federal da Bahia que tenham a responsabilidade sobre a publicação de dados científicos.

As condições para o armazenamento dos dados na plataforma disponibilizada pela universidade são:

- I – preenchimento dos metadados necessários de maneira adequada;
- II – adequação a aspectos éticos e legais envolvidos na questão.

O espaço de armazenamento alocado para cada solicitação aprovada e seu respectivo prazo de manutenção na plataforma levarão em conta a necessidade do proponente e a disponibilidade de recursos.

O custo de armazenamento pelo prazo especificado poderá ser coberto pela UFBA ou demandar aporte de recursos por parte do solicitante elegível, nos termos de portaria específica a ser baixada pelo reitor.

A solicitação para uso da estrutura oferecida pela Universidade de São Paulo deverá ser realizada por meio de preenchimento de formulário disponível no sistema corporativo do Repositório de Dados Científicos da Instituição.

As solicitações serão analisadas por um Grupo de Análise de Dados Científicos, composto por 01 representante de cada um dos seguintes órgãos: a) Pró-Reitoria de Pesquisa e Pós-Graduação (PRPPG); e b) Sistema de Biblioteca da UFBA (Sibi). Todos serão nomeados pelo reitor.

Será de responsabilidade da Superintendência de Tecnologia da Informação (STI) prover recursos tecnológicos de suporte ao Repositório. E em conjunto com os demais órgãos (PRPPG e Sibi) serão definidos as diretrizes e políticas de preservação digital dos conteúdos do Repositório, bem como realizar a curadoria dos metadados, além de criar e fornecer orientações sobre o uso do mesmo.

Como depositar os dados de pesquisas científicas

O Depósito de dados pode ser feito por pesquisadores com vínculo com a UFBA a

partir do seu e-mail de contato, solicitando o acesso (*login*) para o e-mail dos responsáveis (que pode ser a biblioteca ou um grupo criado ou setor destinado para este fim), enviando os seguintes os dados:

- Nome completo;
- E-mail de contato;
- Matrícula UFBA;
- Vínculo com a UFBA (aluno, professor, outro);
- Tipo de arquivo publicado (ex.: tese, dissertação, monografia de especialização/graduação, artigo de periódico) a ser vinculado a estes dados científicos depositados.

Após o recebimento do e-mail, o *login*, a senha e as instruções de acesso ao *software* (base de dados que será criada para este fim) serão enviados na resposta. As orientações para a submissão dos dados de pesquisa podem ser acessadas no tutorial de submissão. Este plano irá ajudar os depositários a criar os Planos de Dados de Pesquisa necessários para a sua pesquisa.

Itens a serem incluídos:

- Descrição dos dados: [Fornecer uma breve descrição das informações a serem coletadas – a natureza, o escopo e a escala dos dados que serão gerados ou coletados.] Estes dados, que serão submetidos ao Setor de Análise, enquadram-se no âmbito do Plano de Dados Abertos, da Instituição.
- Responsabilidade: O pesquisador principal terá a responsabilidade geral pela gestão de dados ao longo do projeto de investigação e irá monitorar o cumprimento do plano.
- Arquivo Designado: Os dados de pesquisa deste projeto serão depositados no repositório digital da UFBA [que será criado para esta finalidade junto com a Informática da Instituição] para garantir que a comunidade de pesquisa tenha acesso a longo prazo aos dados. Para tanto, a UFBA se responsabilizará por treinar uma equipe específica para auxiliar no arquivamento.
- Acesso e Compartilhamento: A UFBA disponibilizará os dados de pesquisa deste projeto para a comunidade de pesquisa em geral. Arquivos de dados de uso público: Esses arquivos, nos quais identificadores diretos e indiretos foram removidos para minimizar o risco de divulgação, podem ser acessados diretamente através do site da UFBA. Depois de concordar com os Termos de Uso, os usuários com uma conta na Instituição e um endereço IP autorizado de uma instituição membro podem baixar os

dados. Arquivos de dados de uso restrito: Esses arquivos são distribuídos nos casos em que a remoção de informações potencialmente identificáveis prejudicaria significativamente o potencial analítico dos dados. Os usuários (e suas instituições) devem solicitar esses arquivos, criar planos de segurança de dados e concordar com outros controles de acesso.

- Seleção e retenção: A UFBA arquivará o conjunto de dados completo e sua documentação a longo prazo, apoiando os dados por meio de tecnologias em constante mudança, novas mídias e formatos de dados.
- Metadados: A UFBA criará metadados necessários à recuperação correta dos dados em conformidade com o padrão (a definir, pode ser o XML) mais relevante para as áreas de conhecimento da Universidade.
- Direitos de Propriedade Intelectual: Os investigadores principais e as suas instituições detêm os direitos de autor dos dados de investigação que geram. Ao depositar na base de dados da UFBA, os investigadores não transferem direitos autorais, mas concedem permissão para que a Instituição reuse os dados e transforme os dados conforme necessário para proteger a confidencialidade do respondente, melhorar a utilidade e facilitar a preservação.
- Ética e Privacidade: Consentimento informado para este projeto, as declarações de consentimento informado, se aplicável, não incluirão linguagem que proíba que os dados sejam compartilhados com a comunidade de pesquisa. Gestão de risco de divulgação: Uma vez depositados, os dados serão submetidos a procedimentos para proteger a confidencialidade de indivíduos cujas informações pessoais possam fazer parte de dados arquivados. Estes incluem: (1) revisão rigorosa para avaliar o risco de divulgação; (2) modificar os dados, se necessário, para proteger a confidencialidade; (3) limitar o acesso a conjuntos de dados em que o risco de divulgação permanece alto e (4) consulta com os produtores de dados para gerenciar o risco de divulgação. A UFBA designará um gerente de dados qualificado certificado em gerenciamento de risco de divulgação para atuar como administrador dos dados enquanto eles estão sendo processados. Os dados serão processados e gerenciados em um ambiente seguro não conectado à rede usando a tecnologia de área de trabalho virtual.
- Formato – Submissão: Os dados e a documentação serão submetidos à UFBA nos formatos recomendados.
- Acesso: A UFBA disponibilizará os arquivos de dados quantitativos em vários formatos amplamente utilizados.
- Preservação: Os dados serão armazenados de acordo com os padrões e práticas da Instituição.
- Arquivamento e preservação: A UFBA aceitará a responsabilidade pela preservação a

longo prazo dos dados de pesquisa após o recebimento de um formulário de depósito assinado. Essa responsabilidade inclui o compromisso de gerenciar iterações sucessivas dos dados se novas ondas ou versões forem depositadas. Ela garantirá que os dados de pesquisa sejam migrados para novos formatos, plataformas e mídias de armazenamento, conforme exigido pelas boas práticas na comunidade de preservação digital. A boa prática para a preservação digital exige que uma organização aborde o planejamento de sucessão para ativos digitais. A universidade compromete-se a designar um sucessor no caso improvável de tal necessidade surgir.

- Armazenamento e *backup*: A UFBA se compromete a colocar uma cópia mestre de cada arquivo digital (ou seja, arquivos de dados de pesquisa, documentação e outros arquivos relacionados) no Armazenamento de Arquivamento da universidade, com várias cópias armazenadas com organizações parceiras em locais designados e sincronizadas com o mestre.

APÊNDICE D - Plano de Gestão de Dados

Plan Overview

A Data Management Plan created using DMPTool

Title: Gestão de dados de pesquisa científica: um estudo de caso na Universidade Federal da Bahia

Creator: Normaci correia dos santos Sena

Affiliation: Federal University of Bahia (ufba.br)

Funder: Digital Curation Centre (dec.ac.uk)

Template: Digital Curation Centre

Project abstract:

O movimento da Ciéncia Aberta traz à tona discussões relevantes acerca da ciéncia colaborativa do séculoo XXI. Neste universo colaborativo, os dados de pesquisas tem sido objeto de estudo em virtude da necessidade social de obter dados de confiança para o avanço científico de forma aberta e gratuita. Além disso, diversas agências de fomento reconhecem a importânciados dados de pesquisa, consequentemente, começaram a exigir uma adequada gestâo destes dados. Isto evidencia a necessidade de assegurar meios e mecanismos para preservação e reutilização dos dados por intermédio de políticas que regulamentem a abertura de dados oriundos de recursos públicos. Neste contexto de pesquisa e enquanto grandes produtoras e consumidoras de dados de pesquisas, as universidades são protagonistas na produção e na disseminação do conhecimento, pois os dados fortalecem as pesquisas e o processo de inovação, ratificam o compromisso social e democrático das instituições públicas. Ao contextualizar a literatura sobre dados de pesquisa, percebe-se uma lacuna nas universidades brasileiras bem como na maioria das agências de fomento na estruturação de políticas que direcionem os pesquisadores à gestâo dos dados científicos já que o movimento da Ciéncia Aberta aponta em suas vertentes um tratamento adequado dos dados de pesquisa de forma a viabilizar o processo de armazenamento, organização, busca, recuperação e difusão dos dados coletados. E isto impacta na adesão (ou não) do pesquisador na abertura e no compartilhamento de sua produção científica, dentre outros motivos. Assim, este estudo tem por objetivo geral:conhecer a percepção dos professores pesquisadores no que tange à gestâo de dados e às ações da UFBA desenvolvidas para a implantação de um plano de gestão de dados de pesquisas no contexto da Ciéncia Aberta. Trata-

se de um estudo de caso único integrado, uma vez que, no Brasil, os estudos a respeito de plano de gestão de dados de pesquisa científica ainda se encontram em estágios embrionários. A saber da importância histórica que é a Universidade Federal da Bahia foi escolhida. Para o percurso metodológico, optou-se como instrumento de coleta de dados o questionário on-line seguindo o modelo da Escala Likert. Por se tratar de uma pesquisa qualitativa, a análise dos dados foi realizada sob a técnica da Lógica Paraconsistente, com a finalidade de tratar e equilibrar os graus de concordância e discordância entre as proposições, evitando falhas interpretativas. Os resultados revelaram uma carência de políticas que norteie as ações da Universidade Federal da Bahia no que se refere à gestão e preservação de dados de pesquisa. Ademais, a desinformação entre os professores a respeito das ações da UFBA que tangenciam o movimento da Ciência Aberta, bem como a prática na inserção no ato de abrir e compartilhar escancara a urgência da universidade de implantação do plano de gestão de dados de pesquisa. Conclui-se, portanto, que os pesquisadores enfrentam inúmeros desafios e problemas quando trabalham com dados de pesquisa e não encontram o apoio institucional para atender às demandas, por exemplo, para a gestão apropriada desses mesmos dados.

Start date: 12-31-2019

End date: 07-30-2023

Last modified: 03-24-2023

Gestão de dados de pesquisa científica: um estudo de caso na Universidade Federal da Bahia

Data Collection

What data will you collect or create?

Serão coletados dados textuais e questionário. Em geral, será realizado levantamento dos cursos de pós-graduação da Universidade Federal da Bahia. Seleção dos cursos com conceito Qualis igual ou superior a 4. Levantamento dos professores permanentes de cada curso de pós-graduação da UFBA com descrição da faixa etária, área de conhecimento e tempo de instituição. Descrição das ações e/ou políticas na Instituição que relate ao movimento da Ciência Aberta. Detalhamento da percepção dos professores em relação a gestão de dados de pesquisa científica, incluindo motivos e necessidades para abertura e compartilhamento dos dados de pesquisa, bem como opiniões a respeito das agências de fomento e repositório de dados, por meio do questionário.

How will the data be collected or created?

Os dados serão coletados por meio de pesquisa bibliográfica, análise do site da Universidade e aplicação de questionário online pela plataforma do *Google Forms*.

Documentation and Metadata

What documentation and metadata will accompany the data?

Serão disponibilizados juntos com os datasets, as planilhas em formato aberto (.ods)

Ethics and Legal Compliance

How will you manage any ethical issues?

Para cumprir as questões éticas e legais sobre a pesquisa, foi solicitado de cada respondente a assinatura (online) do Termo de Consentimento Livre e Esclarecido (TCLE). Os dados obtidos foram anonimizados de maneira que possa garantir a integridade e o sigilo do participante. Os dados serão armazenados em pasta protegida por senha no computador pessoal da pesquisadora e também serão salvos na nuvem do Google Drive. O backup será feito constantemente e serão mantidos até 5

anos após a conclusão da pesquisa. Após esse período, os dados serão excluídos e destruídos definitivamente.

How will you manage copyright and Intellectual Property Rights (IP/IPR) issues?

Os dados coletados durante o estudo serão mantidos pela pesquisadora. Além disso, os termos do Termo de Consentimento Livre e Esclarecido (TCLE) serão claramente expresso que a finalidade dos dados será apenas para esta pesquisa, consequentemente, para os produtos resultantes dela (artigos, capítulos de livros e apresentações em eventos). Os dados serão preparados para compartilhamento, por isso, serão acompanhados de uma licença flexível (Creative Commons CC-BY). Caso haja alguma questão referente aos direitos de propriedade intelectual, esta será sanada em fórum específico.

Storage and Backup

How will the data be stored and backed up during the research?

Os dados serão armazenados em pasta protegida por senha no computador pessoal da pesquisadora e também serão salvos e armazenados em nuvem do Google Drive. O backup será feito constantemente pela pesquisadora que fará também a recuperação dos arquivos.

How will you manage access and security?

O acesso aos dados do computador e dos serviços de armazenamento na nuvem do Google Drive será feito por meio de login e senha de acesso.

Selection and Preservation

Which data are of long-term value and should be retained, shared, and/or preserved?

Todos os dados obtidos por meio do questionário e dos registros oriundos do referencial teórico e observação do site serão mantidos por 5 anos. Após esse período serão destruídos definitivamente.

What is the long-term preservation plan for the dataset?

Os dados preparados e processados serão disponibilizados em formato aberto e em repositórios de dados (Dataverse, por exemplo) logo após a finalização desta pesquisa. Não haverá taxas ou custos adicionais.

Data Sharing

How will you share the data?

Os dados poderão ser usados por pesquisadores que investigam sobre a temática ou semelhante. Os dados serão disponibilizados em formato aberto, em repositórios após a conclusão desta pesquisa. Os dados serão compartilhados através de uma licença flexível CC-BY.

Are any restrictions on data sharing required?

Nenhum dado sensível será disponibilizado. Os dados que serão compartilhados, ao ser usado e reutilizado deverão constar a citação e os créditos de atribuição deverão ser informados. Os dados serão disponibilizados logo após a conclusão desta pesquisa em repositórios de dados. Casos omissos, deverão ser informados a pesquisadora por email: normaci@ufba.br

Responsibilities and Resources

Who will be responsible for data management?

A responsabilidade da implementação deste plano de gestão de dados, bem como das atividades/ações a serem desenvolvidas será da pesquisadora. Este documento será revisado e revisado em todo decurso da pesquisa.

What resources will you require to deliver your plan?

Será de responsabilidade da pesquisadora a seleção e disponibilização dos dados em repositório de dados em acesso aberto e livre. Os softwares que serão utilizados são: para os serviços de armazenamento em nuvem (Google Drive), GitHub (repositório disponível em nuvem e também com uma versão para computador pessoal) e planilhas (Google docs/Excel).

ANEXO A - Plano de dados abertos da UFBA (pda_ufba_2019-2020.pdf)**PDA 2019-2020****Universidade Federal da Bahia
Plano de Dados Abertos****PDA-UFBA
2019-2020****Sumário****1. Apresentação**

Conforme orientação estratégica do Plano de Desenvolvimento Institucional (PDI) 2018-2022 da Universidade Federal da Bahia, o Plano de Dados Abertos (PDA) tem como objetivo orientar as ações de implementação e promoção da iniciativa de abertura de dados da instituição.

A abertura de dados da UFBA promoverá a visibilidade e a transparéncia de informações, atividades, ações e serviços da Universidade, conforme os objetivos estratégicos do PDI UFBA, visando aprimorar e fortalecer os processos e instrumentos de governança institucional. Para além da obrigatoriedade por lei, políticas públicas de informação para conhecimento aberto incentivam a disponibilização de dados para melhoria da gestão pública, o provimento da transparéncia, o estímulo ao controle e participação social e o fomento à inovação tecnológica.

O PDA-UFBA atende as exigências estabelecidas no Decreto nº 8.777, de 11 de maio de 2016, sendo suas ações consonantes com o disposto na Lei de Acesso à Informação (LAI), Lei nº 12.527/2011; Lei Complementar nº 101/2000 (art. 48) e a Instrução Normativa SLTI nº 4, de 13 de abril de 2012 (Infraestrutura Nacional de Dados Abertos). As ações do PDA-UFBA são consonantes também com os compromissos assumidos pelo Governo Federal no âmbito da Parceria para Governo Aberto (Open Government Partnership - OGP).

O PDA da UFBA foi elaborado em 2018 de maneira colaborativa, por uma comissão formada pelos seguintes representantes: 1 (um) representante do Serviço de Informação ao Cidadão (SIC) da UFBA; 1 (um) representante da Ouvidoria Geral da

UFBA; 1 (um) representante da Assessoria de Comunicação Institucional e Imprensa (ASCOM) da UFBA; 1 (um) representante da Pró-Reitoria de Graduação da UFBA (Prograd); 1 (um) representante da Superintendência de Tecnologia da Informação (STI) da UFBA; 1 (um) representante da Assessoria de Tecnologia da Informação da Reitoria da UFBA. Além disso, a comissão foi assessorada por estudante de doutorado da Escola de Administração da UFBA, especialista em dados abertos na administração pública. Após Consulta Pública, com participação aberta da sociedade via Internet, o Plano segue para aprovação pelo Comitê de Governança Digital da UFBA em reunião colegiada de dezembro de 2018.

2. Introdução

A UFBA tem como diretriz estratégica prover transparéncia ativa, sustentabilidade, satisfação e participação do público e comunidade UFBA com a implantação do Plano de Dados Abertos da Universidade.

A UFBA, por meio deste documento, estabelece a sua estratégia para abertura de dados definindo um conjunto de ações de viabilização da prática de dados abertos na instituição - para atuar em conformidade com a Política de Dados Abertos, instituída pelo Governo Federal com a Lei de Acesso à Informação – LAI.

O PDA é um instrumento de planejamento e coordenação da política de disponibilização de dados abertos válido para o biênio 2019-2020 e representa um marco na valorização da transparéncia da administração da UFBA, de forma a permitir o aprimoramento da gestão pública por meio do acesso à informação e do estímulo à participação social.

2.1 Perfil Institucional: Missão, Visão e Valores

A história da Universidade Federal da Bahia confunde-se com os percursos do ensino superior e da pesquisa científica na Bahia há 71 anos - sendo a única Instituição universitária federal no Estado até meados da primeira década dos anos 2000. A Universidade da Bahia foi oficialmente instalada em 2 de julho de 1946 no Terreiro de Jesus, na antiga Faculdade de Medicina, criada em 1808. Desde sua criação até dezembro de 2015, a Instituição levou ao mercado de trabalho 104.157 graduados, 11.747 mestres e 2.955 doutores.

Conforme nossa Constituição Federal, cabe às Universidades, como componentes do Plano Nacional de Educação, contribuir para a melhoria da qualidade do ensino, a formação para o trabalho e a promoção humanística, científica e tecnológico do País (CRFB/1988, art. 214). Nesse sentido, a elaboração e a implantação do PDA são mecanismos pelos quais as universidades públicas contribuem para a consolidação da

democracia, bem como do desenvolvimento humanístico, científico e tecnológico do País.

A Universidade Federal da Bahia deve se firmar e se afirmar como agente capaz de contribuir para a construção de uma sociedade justa, plural e pautada na sustentabilidade. É em torno desse objetivo que são definidos sua missão, visão e valores.

Missão

A Universidade Federal da Bahia, considerando o seu caráter público e a sua grande diversidade interna, tem a missão de produzir e disseminar ciência, tecnologia, arte e cultura, base para a formação sólida de profissionais, docentes e pesquisadores que atuem dentro de elevados padrões de desempenho técnico e ético e sejam cidadãos comprometidos com a democracia e a promoção da justiça social.

Visão

É ser uma Universidade cuja exceléncia da formação seja socialmente reconhecida e cujo modelo de governança e gestão assegure condições para o continuo desenvolvimento científico, tecnológico, artístico e cultural, conciliando uma crescente inserção internacional e forte engajamento no atendimento de demandas sociais, locais e regionais.

Valores

- Indissociabilidade entre Ensino, Pesquisa e Extensão;
- Busca da exceléncia nas suas atividades fins;
- Respeito à diversidade e combate a todas as formas de intolerância e discriminação;
- Abertura e rigor intelectual, criatividade e busca de inovações;
- Valorização das pessoas, diálogo, transparência e estímulo à cooperação;
- Respeito à história e tradição e abertura para a contemporaneidade;
- Rigor ético em suas decisões e ações;
- Busca da equidade no acesso e permanência na Universidade;
- Pluralismo de ideias, promoção de valores democráticos e de cidadania;
- Compromisso com a transformação social;
- Caráter público, gratuito e autônomo da Universidade;
- Sustentabilidade e responsabilidade ambiental.

2.2 Plano de Desenvolvimento Institucional – PDI 2018-2022

O Plano de Desenvolvimento Institucional – PDI 2018-2022 orienta as políticas e práticas de gestão e autogestão da Universidade pelos próximos anos. Tem, portanto, caráter estratégico ao projetar a UFBA que pretendemos construir, sem desconsiderar sua trajetória histórica e os desafios resultantes do contexto em que atua. Trata-se de um documento que busca conciliar os elementos que definem um PDI (Decreto MEC nº. 5.773, de 9 de maio de 2006) com os procedimentos técnicos que embasam o planejamento estratégico de qualquer instituição.

Os Objetivos Estratégicos constituem os grandes alvos que a Universidade Federal da Bahia busca atingir até 2022, em consonância com o Plano Nacional de Educação (PNE) e o Plano Nacional de Pós-Graduação (PNPG), que fixam grandes diretrizes para a educação brasileira, em particular a de nível superior.

As Metas Globais representam os resultados almejados ou a transformação pretendida no desempenho da Universidade ao cabo do período de vigência do PDI e expressam, de forma clara e verificável, como será avaliado, se o objetivo estratégico foi atingido. O processo de planejamento estratégico definiu um conjunto de 10 Objetivos Estratégicos (OE) que se desdobram em Metas Globais.

No eixo dos recursos, o objetivo estratégico OE6 visa “aprimorar e fortalecer os processos e instrumentos de governança institucional, com o fomento à integração, articulação e sinergia das atividades de ensino, pesquisa e extensão e o estímulo à participação cidadã, com transparéncia pública e diálogo com a comunidade universitária e a sociedade”. Para tanto, temos como Diretriz Estratégica “prover transparéncia ativa, sustentabilidade, satisfação e participação do público e comunidade UFBA com a implantação do Plano de Dados Abertos na Universidade e Tramitação Eletrônica de Processos Institucionais”.

Neste contexto, atuamos em conformidade com a Política de Dados Abertos, instituída pelo Governo federal, e com a LAI – Lei de Acesso à Informação. A externalização dos dados da Universidade valoriza a transparéncia, participação ativa, controle social, e retroalimentação de insumos para o alcance de benefícios para a sociedade e academia. A Diretriz está em conformidade com o PEN (Processo Eletrônico Nacional), programa do Governo federal que, através da tramitação eletrônica de processos, visa agilidade, produtividade, satisfação do público usuário, redução de custos, sustentabilidade, compartilhamento de conhecimento, transparéncia ativa, mudança organizacional e qualidade de vida do servidor.

3. Objetivos do Plano de Dados Abertos

3.1. Objetivo Geral

Estabelecer a abertura de dados da Universidade Federal da Bahia ampliando a transparéncia ativa na Administração Pública e na Ciência Aberta, com o compromisso de divulgação permanente da dados públicos em formato aberto.

3.2. Objetivos Específicos

- Identificar prioridades e disponibilizar dados em formatos abertos, de modo a incrementar o processo de transparéncia e de acesso a informações públicas;
- Fomentar a produção de conhecimento, o controle social e a gestão pública participativa, a partir da utilização dos dados abertos pela sociedade;
- Melhorar a gestão da informação estimulando a inovação aberta e o desenvolvimento de soluções tecnológicas baseadas no uso dos dados publicados.

4. Dados para Abertura

Em conformidade com a Política de Dados Abertos do Poder Executivo federal, serão apresentados os critérios usados para definição e priorização dos dados da UFBA que devem ser abertos, garantindo os princípios da publicidade e da transparéncia na administração pública. Dessa forma, assume-se o compromisso de divulgação permanente dos dados de interesse público produzidos nas diversas ações realizadas por esta instituição, considerando-se os princípios da legalidade, da economicidade e da eficiência.

Para se estabelecerem as metas de abertura de dados e as etapas do processo, são consideradas as seguintes diretrizes: (1) Dados relacionados aos pedidos de informação já recebidos pela UFBA, através do Serviço de Informação ao Cidadão (SIC); (2) Dados armazenados nos sistemas de informação usados na UFBA, dos quais serão priorizados os já apresentados nas áreas públicas dos sistemas e também aqueles relacionados às funcionalidades mais acessadas nos sistemas; (3) Informações já publicados no Portal da Transparéncia da instituição; (4) Cumprimento dos normativos legais e compromissos com o Decreto 8.777 de 2016, as normas definidas pela Infraestrutura Nacional de Dados Abertos (INDA) e catalogação dos dados no Portal Brasileiro de Dados Abertos.

Os dados abertos disponibilizados pela UFBA serão coletados, prioritariamente, a partir de informações consolidadas nos sistemas de informação da instituição. Os dados publicados serão periodicamente atualizados, preferencialmente de forma automática.

4.1. Estratégias para Abertura

A abertura de dados se dará através do seguinte roteiro:

- Identificação das categorias de dados candidatas à publicação de acordo com as diretrizes estabelecidas neste PDA;
- Levantamento do conjunto de dados candidato à abertura;
- Priorização e seleção dos dados que serão abertos;
- Definição de responsáveis pelo preparo, pela atualização dos dados e pelo detalhamento ou pela atualização do plano de ação com metas e prazos;
- Consolidação da matriz de responsabilidades e definição da governança e do fluxo de aprovação do PDA e das revisões;
- Utilização de metodologia de abertura de dados a ser seguida pelas áreas responsáveis (padrões INDA e INDE);
- Definição da infraestrutura e da arquitetura tecnológica para abertura dos dados dos sistemas;
- Desenvolvimento e lançamento do portal de dados abertos próprio (Portal de Dados Abertos UFBA), devidamente integrado com os sistemas de informação usados na instituição, acessíveis a partir do endereço <http://dados.ufba.br>

4.1.1 - Proteção de dados pessoais: São considerados dados pessoais informações como endereço, filiação, data de nascimento, estado civil, número de documentos como CPF e RG, nome e/ou quantidade de dependentes. Também são considerados dados pessoais informações sobre saúde e desempenho acadêmico. Os dados disponibilizados são de responsabilidade de cada uma das fontes. Os dados abertos devem estar em conformidade com a LGPD - Lei Geral de Proteção dos Dados Pessoais, Lei 13.709/18, de 14/08/2018.

4.1.2 - Inventário de Dados Abertos da UFBA. O elenco de dados deve contemplar: (1) Grupo do dado, (2) Categoria específica do dado, (3) Unidade responsável, (4) Períocidade de atualização, e (5) Meta para abertura do dado. Será confeccionado inicialmente pela comissão de dados abertos instituída nesse PDA e, em seguida, o inventário passará a ser atualizado e estendido pelos gestores das áreas responsáveis, em cooperação setores da Administração.

5. Catalogação no Portal Brasileiro de Dados Abertos

Os diversos conjuntos de dados abertos produzidos na UFBA serão publicados no Portal de Dados Abertos da Instituição e também catalogados no Portal Brasileiro de Dados Abertos. A execução desses processos é de responsabilidade da Superintendência de Tecnologia da Informação (STI) e será prioritariamente realizada de forma automatizada. Quando os dados a serem publicados não estiverem

PDA 2019-2020

armazenados nos sistemas de informação gerenciais da Universidade, eles serão catalogados diretamente por cada uma das áreas responsáveis, com o apoio da Administração Central, para depois serem publicados manualmente.

6. Sustentação

O processo de disponibilização de dados abertos requer uma permanente revisão a fim de realizar o devido acompanhamento e atualização das informações. Na UFBA, esta função será atribuída aos gestores responsáveis pelos dados, sob a supervisão do Comitê de Governança Digital da UFBA, e com o apoio da Pro-reitoria de Planejamento, através do Serviço de Informação ao Cidadão (SIC), da Ouvidoria Geral, da Assessoria de Comunicação Institucional, da Superintendência de Tecnologia da Informação (STI) e da Reitoria da UFBA.

7. Governança

A Comissão de Dados Abertos será responsável pela gestão das informações disponibilizadas.

Participante	Responsabilidade
Comitê de Governança Digital (CGD)	Definição, coordenação e acompanhamento da execução das atividades previstas no PDA. Zelar para que o PDA esteja em conformidade com o planejamento institucional da universidade, dentre eles o PDI e PDTIC.
Pró-reitoria de Planejamento (Proplan)	Acompanhamento e monitoramento da disponibilização dos dados abertos. Interação com as diversas áreas da Universidade visando promoção da abertura dos dados da instituição.
Pró-reitorias, Superintendências e Unidades Administrativas	Indução da publicação de novos conjuntos de dados relativos à sua área. Catalogação, atualização, evolução dos dados relativos à sua área. Verificação da consistência e da qualidade dos dados publicados.

Superintendência de Tecnologia da Informação (STI)	<p>Curadoria (catalogação, atualização, revisão, publicação ou qualquer alteração) dos metadados dos dados disponibilizados através do portal.</p> <p>Desenvolvimento de ferramentas computacionais para a extração de dados dos sistemas institucionais e a sua disponibilização no formato adequado à sua publicação no Portal de Dados Abertos.</p>
---	--

8. Monitoramento e Controle

O monitoramento e o controle das ações elencadas neste PDA serão realizados pela autoridade de monitoramento da Lei de Acesso à Informação e Ouvidoria. Anualmente será publicado relatório, com informações sobre os dados abertos no período, dados com previsão de abertura e estatísticas de consultas realizadas.

A melhoria da qualidade dos dados terá como base os critérios indicados no Modelo de Maturidade de Dados Abertos da INDA e será de responsabilidade do órgão responsável pelos dados, observando o PDA e podendo propor alterações com objetivo de melhoria continua da qualidade e confiabilidade dos dados publicados.

Os dados abertos da UFBA serão disponibilizados no site www.dados.ufba.br. O PDA será divulgado no Portal Eletrônico da Universidade. Todas as publicações, governança e revisões do plano serão noticiadas também por meio do portal. Também serão empreendidas ações de promoção e fomento a fim de promover a institucionalização do Plano de Dados Abertos. As informações sobre governança e revisão do documento serão publicizadas para a comunidade acadêmica e externa, tanto no portal institucional quanto nas redes sociais da Universidade, administrados pela Assessoria de Imprensa.

9. Plano de Ação Efetuado

Atividade	Prazo	Responsável
Realizar Curso/Oficina para Elaboração de Plano de Dados Abertos com gestores das várias áreas fins e meio da UFBA	18 e 19 de Julho de 2018	EAUFBA, Reitoria
Apresentar levantamento de Inventário de Dados UFBA	Setembro/18	Comissão PDA

PDA 2019-2020

Indicar Comissão para Elaborar PDA-UFBA	Setembro/18	CGD, Gabinete Reitoria
Elaboração da 1ª Versão do Plano de Dados Abertos	Outubro/18	Comissão PDA
Apreciação do PDA-UFBA pelo Comitê de Governança Digital, divulgação e encaminhamento para Consulta Pública	Dezembro/18	CGD, Reitoria, ASCOM
Processo de Consulta Pública	15 dias	Comissão PDA
Aprovação do PDA no Comitê de Governança Digital e Conselho Administrativo da UFBA	Dezembro/18	Assessoria Reitoria
Publicação do primeiro conjunto de dados abertos e catalogação no Portal Brasileiro de Dados Abertos	Março 2019	STI e Ouvidoria