

**UNIVERSIDADE DO ESTADO DE SANTA CATARINA – UDESC**  
**CENTRO DE CIÊNCIAS HUMANAS E DA EDUCAÇÃO**  
**PROGRAMA DE PÓS-GRADUAÇÃO EM GESTÃO DA INFORMAÇÃO**  
**MESTRADO PROFISSIONAL EM GESTÃO DE UNIDADES DE INFORMAÇÃO**

**IGOR YURE RAMOS MATOS**

**REPOSITÓRIO DE DADOS: IMPLANTAÇÃO DA GESTÃO DE DADOS NAS**  
**UNIVERSIDADES**

**FLORIANÓPOLIS**

**2022**

**IGOR YURE RAMOS MATOS**

**REPOSITÓRIO DE DADOS: IMPLANTAÇÃO DA GESTÃO DE DADOS NAS  
UNIVERSIDADES**

Dissertação apresentada ao Programa de Pós-Graduação em Gestão da Informação (PPGInfo), do Centro de Ciências Humanas e da Educação (FAED), da Universidade do Estado de Santa Catarina (UDESC), como requisito parcial para a obtenção do grau de mestre em Gestão da Informação.

Orientador: Prof. Dr. Divino Ignácio Ribeiro Júnior.

**FLORIANÓPOLIS**

**2022**

**Ficha catalográfica elaborada pelo programa de geração automática da  
Biblioteca Setorial do FAED/UDESC,  
com os dados fornecidos pelo(a) autor(a)**

Matos, Igor Yure Ramos

Repositório de dados : implantação da gestão de dados nas universidades / Igor Yure Ramos Matos. -- 2022.  
153 p.

Orientador: Divino Ignácio Ribeiro Júnior  
Dissertação (mestrado) -- Universidade do Estado de Santa Catarina, Centro de Ciências Humanas e da Educação, Programa de Pós-Graduação , Florianópolis, 2022.

1. Gestão de dados de pesquisa. 2. Repositório de dados de pesquisa. 3. Serviço de suporte a dados de pesquisa. 4. Política de gestão de dados. 5. Repositórios digitais confiáveis. I. Ribeiro Júnior, Divino Ignácio. II. Universidade do Estado de Santa Catarina, Centro de Ciências Humanas e da Educação, Programa de Pós-Graduação . III. Título.

**IGOR YURE RAMOS MATOS**

**REPOSITÓRIO DE DADOS: IMPLANTAÇÃO DA GESTÃO DE DADOS NAS  
UNIVERSIDADES**

Dissertação apresentada ao Programa de Pós-Graduação em Gestão da Informação (PPGInfo), do Centro de Ciências Humanas e da Educação (FAED), da Universidade do Estado de Santa Catarina (UDESC), como requisito parcial para a obtenção do grau de mestre em Gestão da Informação.

**BANCA EXAMINADORA**

Prof. Dr. Divino Ignácio Ribeiro Júnior  
Universidade do Estado de Santa Catarina (UDESC)

Membros:

Prof. Dr. Júlio da Silva Dias  
Universidade do Estado de Santa Catarina (UDESC)

Prof. Dr. Rene Faustino Gabriel Junior  
Universidade Federal do Rio Grande do Sul (UFRGS)

Florianópolis, 2 de setembro de 2022.

Dedico à minha avó materna, Maria Ramos Damasceno, sem ela ter feito o papel de mãe, após a precoce partida da minha mãe, eu não teria chegado até aqui! A Melanias Ramos Damasceno (*in memorian*) e Jove Moreira Damasceno (*in memorian*), mãe e avô.

## AGRADECIMENTOS

Agradeço ao meu orientador, Prof. Divino Ignácio Ribeiro Júnior, pela oportunidade, atenção, instrução e ensinamento que vai além deste trabalho! Muito obrigado!

Agradeço aos membros da banca de defesa, Prof. Rene Faustino Gabriel Junior e Prof. Júlio da Silva Dias, que aceitaram avaliar e contribuir na melhoria desta pesquisa. E aos professores que se colocaram à disposição Prof. Fabiano Couto Corrêa da Silva e Profa. Ana Maria Pereira. Muito obrigado!

Agradeço à professora da banca de qualificação, Profa. Elaine Rosangela de Oliveira Lucas. Novamente, ao Prof. Rene Faustino, pela contribuição na qualificação, pelo acesso ao Repositório Dados Preservados da RNP, pela disposição em tirar dúvidas, pelos materiais indicados, foram contribuições fundamentais para o desenvolvimento deste estudo. E aos professores, Paula Carina de Araújo e Fabiano Couto Corrêa da Silva, que se colocaram à disposição na banca de qualificação. Muito obrigado!

Minha gratidão à Roberta Moraes de Bem que, lá atrás, possibilitou minha vinda à UFSC, o seu sim me possibilitou chegar até aqui e você não imagina o bem que me fez. Hoje consigo dizer que tive uma nova oportunidade de crescimento pessoal e profissional. Muito obrigado!

Agradeço à Biblioteca Universitária da Universidade Federal de Santa Catarina (BU/UFSC), na pessoa da diretora Gleide Bitencourte José Ordovás, pela licença concedida, assim como, pela amizade e confiança depositadas. Aos colegas de trabalho: Andrea, Crislaine Zurita (também pela contribuição na leitura da política), Hilda Carolina, Guilherme Righetto, Laísa Cristina, Luziane, Thiago, Cátia, Fabrício e Thaís, pelas conversas, ajudas e trocas de conhecimento! Muito obrigado!

Aos professores e aos técnicos do Programa de Pós-Graduação em Gestão da Informação. Aos alunos da turma 2020 (a turma pandêmica!), com quem convivi remotamente na maior parte do tempo, dividindo ideias, conhecimentos, assuntos, angústias e risadas. Em especial: Débora Russiano, Amanda Herzmann, Dayane Dornelles (também pela leitura e formatação deste trabalho!) e Luciana Schmidt. Muito obrigado!

Agradeço aos colegas da Biblioteca Central da Universidade Federal de Mato Grosso (BC/UFMT), a experiência com o repositório institucional me levou indiretamente ao tema

desta pesquisa. Em especial: Jordan Antônio, Nádia Paes, Valquíria Barbieri, Simone Gomes, Rozangela Zelensk, Rosângela Sohn e Mara Regina. Muito obrigado!

Aos amigos e amigas, de perto ou de longe, presenciais ou remotos. Muito obrigado pelas conversas e incentivos... Desculpe não nomear todos aqui, mas eu poderia cometer o equívoco de deixar alguns de fora.

Agradeço ao Cristiano Moura que me acompanhou do quase até a totalidade deste trabalho... (terapia é vida... façam terapia!!!). Muito obrigado!

Agradeço aos meus filhos de quatro patas: Lola, Mileyde, Amoisés, Marcelina, Branquinha, Pretinha e Lord. Pela companhia, pela distração, pela responsabilidade, por pedirem atenção e por me lembrarem de sair da frente do computador. Muito obrigado e que vivam muito!

Agradeço aos familiares (alguns!) que acompanharam de longe o desenvolvimento deste trabalho!!!

Aos bons professores que passaram por minha vida, tanto pelo conteúdo ministrado quanto pelos exemplos e incentivos!!!

A todos que me auxiliaram, desde a etapa de seleção até a conclusão desta pesquisa. Obrigado ao colega Toni Picalho que praticamente me apresentou a UDESC e me incentivou a realizar a seleção. A professora Úrsula Blattmann pelas sugestões ao trabalho. Muito obrigado!

“[...] embora o mundo não mude com a mudança de paradigma, depois dela o cientista trabalha em um mundo diferente” (Thomas Kuhn).



## RESUMO

Os dados de pesquisa são elementos valiosos no contexto da *e-Science*, caracterizada pela intensiva produção e uso de dados, com grande potencial de inovação e de avanços do conhecimento científico, principalmente, com o movimento da ciência aberta. As agências de financiamento internacionais, dos países em desenvolvimento, exigem o depósito dos dados brutos, economizando tempo e recursos nas pesquisas científicas, uma realidade que, aos poucos, começa a ser exigida no Brasil, para isso, é necessário a implantação de repositórios de dados nas universidades brasileiras. A curadoria deverá ficar a cargo dos bibliotecários de dados, sendo necessários novos conhecimentos e habilidades para essa demanda. Com o objetivo de desenvolver um modelo para o projeto e a implantação de repositório de dados, sendo uma pesquisa exploratória e descritiva, com abordagem qualitativa, que utiliza as técnicas de pesquisa bibliográfica, documental e experimental, analisou-se a plataforma Dataverse e as suas funcionalidades. A plataforma cumpre com a função de armazenamento dos conjuntos de dados; da documentação, plano de gestão de dados e glossário de termos; e do versionamento dos *datasets*, atende aos padrões internacionais de metadados, assim, permite o reuso e a possibilidade da preservação a longo prazo com a integração à plataforma Archivematica, possibilita, também, o uso de identificadores persistentes como o Handle e o DOI. Como resultado de pesquisa, realizou-se uma proposta composta por nove fases para a implantação da gestão de dados nas universidades brasileiras, identificando os atores institucionais: Biblioteca Universitária, Pró-Reitoria de Pesquisa, Pró-Reitoria de Pós-Graduação e Secretaria de Tecnologia da Informação, definindo seus papéis e suas responsabilidades. Apresentou-se um Serviço de Suporte a Dados de Pesquisa para as Bibliotecas Universitárias, dividido em: suporte, orientação e curadoria de dados. Baseou-se nos requisitos da certificação CoreTrustSeal para a elaboração de uma minuta de política de gestão de dados e análise da plataforma Dataverse. Percebe-se a necessidade da implantação da gestão de dados nas Instituições de Ensino Superior, sendo que o serviço de suporte a dados de pesquisa nas bibliotecas universitárias permitirá a perenidade de ações, a mudança da cultura dos pesquisadores adotando as boas práticas, tendo o propósito de compartilhar dados reusáveis e interoperáveis, implementando, assim, o funcionamento dos repositórios de dados nas universidades brasileiras.

**Palavras-chave:** Gestão de dados de pesquisa. Repositório de dados de pesquisa. Serviço de suporte a dados de pesquisa. Política de gestão de dados. Repositórios digitais confiáveis.

## ABSTRACT

Research data are valuable elements in the context of e-Science, characterized by the intensive production and use of data, with great potential for innovation and advances in scientific knowledge, especially with the open science movement. International funding agencies from developing countries require the deposit of raw data, saving time and resources in scientific research, a reality that is slowly beginning to be required in Brazil, for which it is necessary to implement data repositories in universities Brazilians. The curatorship should be in charge of data librarians, requiring new knowledge and skills for this demand. With the objective of developing a model for the design and implementation of a data repository, being an exploratory and descriptive research, with a qualitative approach, which uses the techniques of bibliographic, documentary and experimental research. The Dataverse platform and its functionalities were analyzed. The platform fulfills the function of storing data sets; documentation: Data Management Plan and glossary of terms; and dataset versioning, meets international metadata standards, thus allowing reuse and the possibility of long-term preservation with the integration with the Archivematica platform, it also allows the use of persistent identifiers such as Handle and DOI. As a result of the research, a proposal was made consisting of nine phases for the implementation of data management in Brazilian universities, identifying the institutional actors: University Library, Dean of Research, Dean of Graduate Studies and Department of Technology at Information, defining their roles and responsibilities. A Research Data Support Service for University Libraries was presented, divided into: support, guidance and data curation. It was based on the CoreTrustSeal certification requirements for the elaboration of a data management policy draft and analysis of the Dataverse platform. The need to implement data management in Higher Education Institutions is perceived, and the support service for research data in university libraries will allow for the continuity of actions, changing the culture of researchers by adopting good practices, with the purpose of to share reusable and interoperable data, thus implementing the functioning of data repository in Brazilian universities.

**Keywords:** Research data management. Research data repository. Research Data Support Service. Data Management Policy. Trusted Digital Repositories.

## LISTA DE FIGURAS

Figura 1 – Ciclo de dados de pesquisa DATAONE .....	41
Figura 2 – <i>Curation Lifecycle Model</i> .....	43
Figura 3 – <i>DDI Lifecycle</i> .....	45
Figura 4 – Conceito, tipo de unidade e unidade .....	45
Figura 5 – <i>Data Lifecycle Model</i> .....	46
Figura 6 – Página DMPTool.....	54
Figura 7 – Ferramenta DMPTool .....	56
Figura 8 – Ferramenta DMPonline.....	57
Figura 9 – Fluxo da certificação CoreTrustSeal.....	73
Figura 10 – Geolocalização dos repositórios certificados com a CoreTrustSeal ou anteriores	76
Figura 11 – Estrutura do <i>software</i> Dataverse .....	88
Figura 12 – Representação da estrutura hierárquica do Dataverse.....	89
Figura 13 – Permissões dos perfis Dataverse .....	90
Figura 14 – Criando uma conta Dataverse .....	93
Figura 15 – Conectar ao Dataverse.....	93
Figura 16 – Opções para configuração .....	94
Figura 17 – Como adicionar DATAVERSE e conjuntos de dados.....	95
Figura 18 – Criando um novo Dataverse.....	96
Figura 19 – Hierarquia dos dataverses criados.....	97
Figura 20 – Dataverse novo.....	97
Figura 21 – Como atribuir permissões .....	98
Figura 22 – Atribuindo funções e permissões no Dataverse .....	99
Figura 23 – Sobre o livro de visitas de um conjunto de dados.....	100
Figura 24 – Token de API .....	102
Figura 25 – Token de API criado .....	102
Figura 26 – Submissão de conjunto de dados .....	104
Figura 27 – Implantação da gestão de dados nas universidades.....	108
Figura 28 – Atores envolvidos na implementação da gestão de dados .....	109
Figura 29 – Orientação para dados .....	114
Figura 30 – Suporte a dados .....	115
Figura 31 – Curadoria de dados.....	115
Figura 32 – Serviço de suporte a dados.....	116

## LISTA DE QUADROS

Quadro 1 – Fonte dos dados de pesquisa.....	33
Quadro 2 – Tipologia dos dados.....	34
Quadro 3 – Guia dos princípios FAIR.....	36
Quadro 4 – Etapas do <i>Curation Lifecycle Model</i> .....	43
Quadro 5 – Etapas do <i>Research DATA LIFECYCLE</i> .....	46
Quadro 6 – Atividades típicas realizadas no ciclo de vida dos dados de pesquisa .....	47
Quadro 7 – Elementos principais e elementos adicionais de citação DMPTool.....	51
Quadro 8 – Informações sobre dados de pesquisa.....	52
Quadro 9 – Definições dos metadados do DMPTool.....	55
Quadro 10 – Estrutura de metadados das ferramentas DMPToll e DMPonline.....	58
Quadro 11 – Licenças <i>Creative Commons</i> .....	60
Quadro 12 – Licenças <i>Open Data Commons</i> .....	63
Quadro 13 – Elementos de uma política para gestão de dados de pesquisa.....	69
Quadro 14 – Nível de conformidade para cada um dos requisitos.....	74
Quadro 15 – Critérios CoreTrustSeal .....	74
Quadro 16 – <i>Software</i> e ferramentas desenvolvidas pelo FAIRsFAIR.....	79
Quadro 17 – Lista de termos para pesquisa em bases de dados .....	81
Quadro 18 – Fontes documentais utilizadas .....	83
Quadro 19 – Desenho da pesquisa.....	84
Quadro 20 – Modelo de citação Dataverse.....	100
Quadro 21 – Campos de citação de datasets no Dataverse.....	100
Quadro 22 – Ferramentas externas .....	103
Quadro 23 – Responsabilidade dos atores .....	109
Quadro 24 – Responsabilidades dos pesquisadores .....	110
Quadro 25 – Responsabilidade dos bibliotecários.....	111

## LISTA DE ABREVIATURAS E SIGLAS

API	Application Programming Interface
BN	Biblioteca Nacional
BRAPCI	Base de Dados Referenciais de Artigos de Periódicos em Ciência da Informação
CAPES	Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
CC	Creative Commons
CI	Ciência da Informação
CNPq	Conselho Nacional de Desenvolvimento Científico e Tecnológico
CNPTIA/Embrapa	Embrapa Informática Agropecuária
CRNI	Corporation for National Research Initiatives
CSV	Comma Separated Values
CVDP	Ciclo de vida dos dados de pesquisa
DATAONE	Data Observation Network for Earth
DCC	Digital Curation Centre
DDI ALIANCE	Data Documentation Initiative Alliance
DMP	Data Management Plan
DOI	Digital Object Identifier
DSA	Data Seal of Approval
DANS	Data Archiving and Networked Services
DC	Dublin Core
WDS	World Data System
ICS	International Science Council
EI	Encontrabilidade da Informação
EMBRAPA	Empresa Brasileira de Pesquisa Agropecuária
FAIR	Findable, Accessible, Interoperable, Reusable
FAIRsFAIR	Fostering Fair Data Practices in Europe
FAPESP	Fundação de Amparo à Pesquisa do Estado de São Paulo
FTP	File Transfer Protocol
FURG	Universidade Federal do Rio Grande
IBGE	Instituto Brasileiro de Geografia e Estatística
IBICT	Instituto Brasileiro Informação em Ciência e Tecnologia
IES	Instituição de Ensino Superior
ITA	Instituto Tecnológico de Aeronáutica
IQSS	Institute for Quantitative Social Science
ISSO	International Organization for Standardization
LISA	Library and Information Science Abstracts
MAA	Movimento de Acesso Aberto
MIT	Instituto de Tecnologia de Massachusetts
NESTOR	Network of Expertise in Long-Term Storage of Digital Resources

NSF	National Science Foundation
OAI-PMH	Open Archives Initiative Protocol for Metadata Harvesting
ODC	Open Data Commons
OKF	Open Knowledge Foundation
ODC-BY	Open Data Commons Attribution License
ODbL	Open Data Commons Open Database License
PDDL	Public Domain Dedication and License
P&D	Pesquisa e Desenvolvimento
PGD	Plano de Gestão de Dados
PPGInfo	Programa de Pós-Graduação em Gestão da Informação
Rede Cariniana	Rede Brasileira de Serviços de Preservação Digital
RDF	Resource Description Framework
RDP BRASIL	Grupo de Trabalho de Redes de Dados de Pesquisa Brasileira
RE3DATA	Registry of Research Data Repositories
Rede Cariniana	Rede Brasileira de Serviços de Preservação Digital
RNP	Rede Nacional de Ensino e Pesquisa
SciELO	Scientific Electronic Library Online
TICs	Tecnologias de Informação e Comunicação
UDESC	Universidade do Estado de Santa Catarina
EU	União Europeia
UFRGS	Universidade Federal do Rio Grande do Sul
USP	Universidade de São Paulo
VDC	Virtual Data Center
XML	Extensible Markup Language

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO.....</b>	<b>14</b>
1.1	PROBLEMA DE PESQUISA .....	16
1.2	OBJETIVO GERAL .....	16
1.3	OBJETIVOS ESPECÍFICOS .....	16
1.4	JUSTIFICATIVA .....	17
1.5	ESTRUTURA DO TRABALHO .....	18
<b>2</b>	<b>FUNDAMENTAÇÃO TEÓRICA.....</b>	<b>21</b>
2.1	A ATIVIDADE CIENTÍFICA .....	21
<b>2.1.1</b>	<b>A Ciência da Informação no uso dos dados.....</b>	<b>25</b>
<b>2.1.2</b>	<b>e-Science.....</b>	<b>28</b>
2.2	DADOS DE PESQUISA .....	30
<b>2.2.1</b>	<b>Os princípios FAIR.....</b>	<b>35</b>
2.3	A GESTÃO DE DADOS DE PESQUISA .....	39
<b>2.3.1</b>	<b>Ciclo de vida dos dados de pesquisa (CVDP).....</b>	<b>40</b>
2.3.1.1	<i>Ciclo de vida de Dados DATAONE.....</i>	<i>41</i>
2.3.1.2	<i>CURATION LIFECYCLE MODEL .....</i>	<i>42</i>
2.3.1.3	<i>DDI LIFECYCLE MODEL.....</i>	<i>44</i>
2.3.1.4	<i>Research Data Lifecycle .....</i>	<i>46</i>
<b>2.3.2</b>	<b>Descrição dos metadados .....</b>	<b>48</b>
<b>2.3.3</b>	<b>Elementos para a citação do conjunto de dados .....</b>	<b>50</b>
<b>2.3.4</b>	<b>O Plano de Gestão de Dados (PGD).....</b>	<b>51</b>
2.3.4.1	<i>A ferramenta DMPTool.....</i>	<i>54</i>
2.3.4.2	<i>A ferramenta DMPonline .....</i>	<i>56</i>
<b>2.3.5</b>	<b>Tipos de licenças .....</b>	<b>59</b>
2.3.5.1	<i>As licenças Creative Commons .....</i>	<i>60</i>
2.3.5.2	<i>Open Data Commons.....</i>	<i>61</i>
2.4	O SERVIÇO REPOSITÓRIO DE DADOS: ANTECEDENTES .....	64
2.5	REPOSITÓRIO DE DADOS .....	65
2.6	A INSTITUCIONALIZAÇÃO DO REPOSITÓRIO DE DADOS.....	67
<b>2.6.1</b>	<b>Políticas para repositórios de dados.....</b>	<b>68</b>
2.7	SERVIÇO DE SUPORTE À GESTÃO DE DADOS DE PESQUISA NAS UNIVERSIDADES .....	70

2.8	CERTIFICAÇÃO CoreTrustSeal .....	71
<b>2.8.1</b>	<b>O projeto FAIRsFAIR: em busca da certificação Core Trust Seal .....</b>	<b>76</b>
<b>3</b>	<b>PROCEDIMENTOS METODOLÓGICOS.....</b>	<b>80</b>
3.1	CARACTERIZAÇÃO DA PESQUISA .....	80
3.2	QUANTO AOS PROCEDIMENTOS .....	80
<b>3.2.1</b>	<b>Pesquisa bibliográfica.....</b>	<b>80</b>
<b>3.2.2</b>	<b>Pesquisa documental .....</b>	<b>82</b>
<b>3.2.3</b>	<b>Pesquisa experimental.....</b>	<b>84</b>
3.3	DESENHO DA PESQUISA.....	84
<b>4</b>	<b>ANÁLISE DA PLATAFORMA DATAVERSE .....</b>	<b>86</b>
4.1	A PLATAFORMA DATAVERSE.....	86
<b>4.1.1</b>	<b>Perfis da plataforma DATAVERSE .....</b>	<b>89</b>
<b>4.1.2</b>	<b>Metadados do Dataverse .....</b>	<b>90</b>
<b>4.1.3</b>	<b>Os identificadores persistentes .....</b>	<b>91</b>
4.2	TESTE NA PLATAFORMA DATAVERSE.....	92
<b>5</b>	<b>MODELO PARA A IMPLANTAÇÃO DE REPOSITÓRIO DE DADOS DE PESQUISA .....</b>	<b>106</b>
5.1	A IMPLANTAÇÃO DA GESTÃO DE DADOS NAS UNIVERSIDADES.....	107
5.2	OS ATORES ENVOLVIDOS .....	109
5.3	GRUPO DE TRABALHO DE GESTÃO DE DADOS .....	110
5.4	AS RESPONSABILIDADES DOS PESQUISADORES E BIBLIOTECÁRIOS .....	110
5.5	POLÍTICA DE GESTÃO DE DADOS .....	112
5.6	SERVIÇO DE GESTÃO DE DADOS DE PESQUISA.....	113
<b>6</b>	<b>CONSIDERAÇÕES FINAIS.....</b>	<b>117</b>
	<b>REFERÊNCIAS .....</b>	<b>120</b>
	<b>APÊNDICE A – CICLO DE VIDA DATAONE .....</b>	<b>129</b>
	<b>APÊNDICE B – CRITÉRIOS CORE TRUST SEAL.....</b>	<b>134</b>
	<b>APÊNDICE C – CAMPOS DO DMPONLINE .....</b>	<b>141</b>
	<b>APÊNDICE D – POLÍTICA DE GESTÃO DE DADOS DE PESQUISA .....</b>	<b>145</b>
	<b>ANEXO A – PADRÕES DE METADADOS .....</b>	<b>152</b>



## 1 INTRODUÇÃO

A ciência no século XXI tem realizado avanços, entre outros fatores, devido às tecnologias de informação e modernos computadores, assim, percebe-se a urgência do fazer científico, por isso, cooperar é a melhor solução, além de diminuir os custos operacionais com esse trabalho colaborativo. Um bom exemplo dessa cooperação é o reuso dos dados de pesquisa, que são reutilizados por outros pesquisadores e centros de pesquisa, além dos trabalhos de coleta que podem ser feitos por diferentes grupos de pesquisas e armazenados em conjunto em um mesmo *dataset*<sup>1</sup>. Os avanços da ciência e as mudanças no mundo científico são estudados na Ciência da Informação (CI), na área da comunicação científica.

A percepção da necessidade de disponibilizarem as pesquisas e torná-las acessíveis é algo recente para os cientistas, ideias que foram sendo disseminadas e praticadas com o Movimento de Acesso Aberto<sup>2</sup> (MAA), um exemplo de iniciativa brasileira é a Scientific Electronic Library Online (SciELO), em 1997. As mudanças de prática dos pesquisadores demandam grandes esforços e as exigências geralmente partem de algum órgão superior, como exemplo, as agências de fomento que realizam os financiamentos das pesquisas.

Desde seu surgimento, o MAA obteve avanços significativos e, em conjunto com o Movimento de Software Livre, permitiu o desenvolvimento de bibliotecas digitais de teses e dissertações, repositórios digitais, revistas online, portais de periódicos, entre outros. Os próximos passos são disponibilizarem na íntegra tudo o que se produz, ou seja, também os dados brutos que subsidiam as publicações científicas. Segundo Silva (2019, p. 3), os dados são valiosos porque, “[...] se o conhecimento é o motor do avanço científico, os dados são seu combustível”.

As evoluções ocorridas com computadores e equipamentos de alta performance permitem, hoje, grandes capturas e transferências de dados, tanto para pesquisadores e universidades, quanto para empresas e indivíduos da sociedade civil. Grandes computadores, telescópios e outros equipamentos conseguem capturar imensas quantidades de dados, sendo esses tratados como o elemento principal para o desenvolvimento de pesquisas acadêmicas, os cientistas chamam de *e-Science*, este momento de obter grandes conjuntos de dados por meio de equipamentos tecnológicos.

---

<sup>1</sup> *Dataset* é um conjunto de dados disponíveis em um ou mais arquivos digitais. Podem ter formatos diferentes como texto, imagens, vídeos, áudio etc. São os objetos que serão objetos de depósito nos repositórios de dados.

<sup>2</sup> O Movimento de Acesso Aberto (MAA) foi iniciado nos anos 2000 com a finalidade de abertura de revistas científicas. Já o Acesso Aberto é um movimento secundário que dá continuidade ao MAA e o amplia com mais iniciativas como, por exemplo, os dados de pesquisa, a ciência cidadã, os dados abertos, entre outras mais recentes.

O fazer científico no contexto da *e-Science* é pensado no reuso, no compartilhamento ético, no trabalho em parceria e no rápido desenvolvimento, ou seja, no progresso científico colaborativo. O reuso dos dados permite a realização de pesquisas com eficiência e rapidez, contribuindo na construção de novos conhecimentos científicos. As informações publicadas em artigos científicos, teses ou dissertações são somente a ponta do *iceberg* (GRAY, 2011), ficando a grande parte dos dados desconhecidos e/ou perdidos dentro das mídias de armazenamento. São esses os dados que precisam ser tratados, depositados e preservados.

Os dados de pesquisa são os elementos coletados para comprovarem hipóteses de pesquisa. Eles subsidiam as publicações primárias, ou seja, são os elementos básicos que comprovam a teoria utilizada na revisão de literatura. Para Silva (2019, p. 21), “[...] os dados científicos são todas as evidências que um investigador necessita para validar suas conclusões após uma pesquisa”. Os dados estão presentes em todas as áreas do conhecimento, são coletados por diferentes instrumentos e a sua grandeza pode variar bastante. Segundo Silva (2019, p. 6), “As possibilidades de reutilização dos dados se ampliam cada vez mais. No que se refere à volume, na atualidade estamos escalando desde *terabytes* a *zettabytes*”.

Para o compartilhamento de dados de pesquisa é necessário a implantação de plataformas específicas para depósito. Os repositórios de dados são uma dessas plataformas tecnológicas que auxiliam na gestão, no compartilhamento e no reuso de dados de pesquisa. Neste contexto, a plataforma Dataverse é um *software* desenvolvido desde 2006 pelo The Dataverse Project, do Institute for Quantitative Social Science (IQSS), da Harvard University, e coordenado pelo pesquisador Gary King.

O Dataverse tende a ser a plataforma de repositório de dados mais utilizada no Brasil, devido a uma aliança realizada em 2015 com a Universidade de Harvard e o Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT), que coordena a Rede Brasileira de Preservação Digital (Rede Cariniana). Segundo Gabriel Júnior *et al.* (2021, p. 6), “[...] o Projeto Dataverse é um aplicativo da web de código aberto para compartilhar, preservar, citar, explorar e analisar dados de pesquisa”.

Para a implantação de um repositório de dados eficiente é necessário a gestão dos dados, ou seja, todo o planejamento e as mudanças de práticas dos pesquisadores para a gestão desses dados para que se tornem acessíveis, encontráveis, interoperáveis e reusáveis, ou seja, atendam aos princípios FAIR. O Plano de Gestão de Dados (PGD) é o documento que descreve: a coleta dos dados, a descrição dos metadados, as licenças atribuídas, os formatos dos arquivos, os identificadores permanentes, a preservação a longo prazo, ou seja, todo o tratamento do ciclo de vida dos dados.

Os bibliotecários são os profissionais que realizam a gestão dos repositórios digitais, logo, os repositórios de dados deverão ficar a cargo das bibliotecas e de seus profissionais (ALMEIDA, 2019). Sendo assim, é fundamental que os bibliotecários se capacitem para essa nova demanda. De acordo com Semeller (2017, p. 18), “[...] os repositórios digitais de dados de pesquisa são a parte da ciberinfraestrutura de *e-Science* que deve ser investigada pelos bibliotecários de dados”. Neste contexto, encontra-se o papel do bibliotecário enquanto curador de dados de pesquisa.

Percebe-se que novas demandas são exigidas para a curadoria de dados, com a principal finalidade de compartilhamento e de reuso, que deverá ser implementado por uma plataforma de repositório de dados, com uma equipe qualificada para gestão dos serviços do repositório. Assim, busca-se criar um modelo para que as universidades implementem a governança de dados.

## 1.1 PROBLEMA DE PESQUISA

No contexto das instituições de pesquisa e universidades, as bibliotecas utilizam diversos recursos para realizar a gestão de acervos de documentos físicos e digitais e, normalmente, dispõe de recursos de *software*, de processos e de serviços bem-estabelecidos para lidar com eles no seu cotidiano. No entanto, o mesmo não ocorre quando se trata da gestão de dados de pesquisa, conforme se observa na literatura supramencionada. Daí decorre a caracterização de uma problemática comum no âmbito dessas instituições, que pode ser definida como a dificuldade de implantação da gestão de dados nas bibliotecas universitárias.

## 1.2 OBJETIVO GERAL

Propor um modelo para a implantação da gestão de dados de pesquisa nas universidades brasileiras.

## 1.3 OBJETIVOS ESPECÍFICOS

- a) Levantar os fundamentos teóricos e recomendações relacionadas à gestão de dados e à implantação de repositórios de dados de pesquisa;
- b) Modelar o serviço de suporte à gestão de dados de pesquisa para bibliotecas universitárias;
- c) Analisar o Dataverse enquanto plataforma para repositórios de dados de pesquisa;

- d) Propor minuta de uma política de gestão de dados que contemple as boas práticas de gestão e de preservação de dados.

#### 1.4 JUSTIFICATIVA

As práticas de *e-Science* são pensadas para uma maior integração da ciência e para trocas de conhecimento entre os pesquisadores e instituições, com a otimização dos custos da pesquisa, criando um fazer colaborativo. O compartilhamento dos dados de pesquisa é uma das práticas de *e-Science*, cujas vantagens são: a) economia de tempo dos pesquisadores; b) economia dos gastos com instrumentos e outros equipamentos de laboratórios, *softwares* e tecnologias específicas, como: computadores, câmeras, sensores, drones, microscópios, lupas, aparelhos de ressonância, *scanners*, sondas e inúmeros dispositivos com alto valor comercial; c) evita-se a duplicação da coleta de dados; e d) pesquisadores são citados e reconhecidos quando seus dados são reutilizados (GRAY, 2011; SEMELER, 2017; SILVA, 2019). Como resultado, por exemplo, acelera a descoberta da cura de doenças, pois as equipes de pesquisadores são potencializadas/diversificadas e, assim, contribuem com o aumento das pesquisas em países em desenvolvimento.

A curadoria dos repositórios de dados de pesquisa deverá ficar a cargo dos “bibliotecários de dados” (SEMELER, 2017). A área da biblioteconomia é chamada a desenvolver novas competências de seus profissionais para trabalharem com essas novas práticas do fazer científico.

Segundo Mueller (2007), a preocupação com a informação científica está na origem da CI, e os estudos relacionados com a comunicação científica continuam centrais à área e também às questões práticas de como prover acesso e preservar o conhecimento. A ciência aberta é uma realidade crescente e, cada dia mais, aplicada à realidade das bibliotecas universitárias brasileiras.

Do ponto de vista pessoal, acredita-se que a Biblioteconomia se renovará quando seus profissionais conseguirem utilizar mais das tecnologias e das inovações em acesso aberto. A curadoria de dados de pesquisa é uma reinvenção das atividades técnicas do profissional bibliotecário, pois, esse é chamado a reinventar-se no trabalho com outros suportes além dos bibliográficos. Assim, as novas práticas de *e-Science* precisam ser incorporadas à prática biblioteconômica.

Sendo o autor um bibliotecário que atua em uma biblioteca universitária, sabe que este tema ainda não faz parte da maioria dessas bibliotecas, de acordo com Oliveira e Silva (2016,

p. 6), “[...] a realidade brasileira voltada para a ciência aberta e dados de pesquisa encontra-se em um estágio incipiente”, assim, esta pesquisa tem a chance de ser um produto de aprendizado para os colegas de profissão.

A contribuição para a linha de pesquisa Gestão de Unidades de Informação, do Programa de Pós-Graduação em Gestão da Informação (PPGInfo), da Universidade do Estado de Santa Catarina (UDESC), se justifica por ser um tema em ascensão, com grande demanda por parte das universidades brasileiras às suas bibliotecas e, assim, há exigência aos seus bibliotecários que estão sendo instigados a aprenderem, trabalharem e se capacitarem em curadoria e gestão de dados de pesquisa, mas também a participarem da escolha e da instalação dos *softwares* de repositórios de dados, ou seja, eles necessitarão de materiais e de literatura para desenvolverem as competências e os conhecimentos para o trabalho de curadoria dos dados.

Os trabalhos dessa linha de pesquisa giram em torno de “estudos acerca da implementação de processos e do uso de tecnologias de informação e comunicação em unidades de informação” (UDESC, 2021), sendo que, a implantação/instalação de *softwares* e de outras tecnologias é algo distante da realidade da maioria dos bibliotecários. Se esses profissionais tiverem uma pesquisa que consiga responder a eles sobre a instalação de *software* de repositório de dados, os tipos de licenças abertas, como trabalhar com os metadados, os identificadores persistentes, a citação dos dados, os modelos de um Plano de Gestão de Dados (PGD), essa pesquisa terá cumprido o seu papel social e trará boa contribuição à área e ao programa, principalmente, para um mestrado profissional.

Portanto, a *e-Science* é uma realidade para os pesquisadores, as universidades e os centros de pesquisa de todo o mundo. Os dados de pesquisa são os elementos centrais para o desenvolvimento da ciência. O compartilhamento desses dados é essencial para a verificação de pesquisas (seja para refazê-las ou testá-las). A cooperação pode interligar importantes pesquisadores ao redor do mundo em um único estudo, pode fortalecer e dar consistência, credibilidade e relevância para as universidades e os centros de pesquisa de países em desenvolvimento, como é o caso do Brasil.

## 1.5 ESTRUTURA DO TRABALHO

O segundo capítulo, denominado como fundamentação teórica, apresenta a atividade científica, o conceito de ciência e de conhecimento científico, a comunidade científica, os estudos de comunicação científica na ciência da informação e a *e-Science* enquanto o quarto paradigma científico. Na seção 2.2, expõem-se os dados de pesquisa e os princípios FAIR. Na

seção 2.3, conta-se com a gestão de dados de pesquisa para contextualizar o ciclo de vida dos dados de pesquisa e exemplificá-los, com quatro modelos de ciclos diferentes, na sequência, aborda-se a importância da descrição dos metadados, os elementos para a citação de um conjunto de dados, o que é o Plano de Gestão de Dados (PGD) e as ferramentas DMPonline e DMPTool para a construção desses e, por fim, comenta-se sobre os tipos de licenças abertas que devem ser atribuídas para conjuntos de dados, sendo as licenças *Creative Commons* ou *Open Data Commons*.

Na seção 2.4, apresentam-se os antecedentes do serviço de repositório de dados, em seguida, na seção 2.5, os repositórios de dados em si. A seção 2.6 expõe a institucionalização do repositório de dados e na subseção 2.6.1 as políticas de repositórios de dados. Na seção 2.7, o serviço de suporte a gestão de dados para as universidades brasileiras. Por fim, a seção 2.8 versa sobre a certificação CoreTrustSeal com diretrizes para que os repositórios de dados se tornem um serviço com qualidade, garantia de confiança e preservação dos dados nele depositados, relata, ainda, O projeto FAIRsFAIR: em busca da certificação Core Trust Seal, que incentivou diversos repositórios de dados europeus a obterem a certificação.

O terceiro capítulo apresenta os procedimentos metodológicos adotados no trabalho, a caracterização da pesquisa, os procedimentos utilizados e as técnicas usadas conforme cada objetivo específico. De acordo com a caracterização é classificada como pesquisa exploratória e descritiva e quanto aos procedimentos como bibliográfica, documental e experimental.

No quarto capítulo apresenta a análise da plataforma Dataverse, os responsáveis pelo seu desenvolvimento desta e a parceria com o IBICT no Brasil, e subseções sobre os usuários da plataforma Dataverse, sobre os Metadados do Dataverse, Os identificadores persistentes adotados pela plataforma e as capturas de tela dos testes realizados nesta.

O quinto capítulo apresenta um fluxo com nove fases, como guia para a implantação da gestão de dados e dos repositórios de dados nas universidades brasileiras, os atores institucionais (Biblioteca, Pró-Reitoria de Pesquisa, Pró-Reitoria de Pós-Graduação e Secretária da Tecnologia da Informação) e suas responsabilidades, a formação de um grupo de trabalho, as responsabilidades dos bibliotecários e dos pesquisadores para que aconteça a gestão de dados, elaborou-se uma minuta de política e apresentou-se um serviço de gestão de dados de pesquisa para as bibliotecas.

No sexto capítulo, constam as considerações finais, nelas apresenta-se o que foi realizado nesta pesquisa de acordo com os objetivos do estudo e, ainda, aborda-se outras propostas de estudo sobre gestão de dados, repositórios digitais confiáveis (certificados) e autoria dos dados de pesquisa.

Nesta dissertação, adotar-se-á o termo “dados de pesquisa” para definir os arquivos que são depositados em repositórios de dados. Os dados de pesquisa englobam um contexto maior, como: dados de pesquisa de senso, dados governamentais, dados de instrumentos, satélites, sensores etc. Os autores Silva (2019), Ribeiro *et al.* (2010), Monteiro (2017), Monteiro, Sant’Ana (2019), Oliveira e Silva (2016), Machado (2015), USP (2020), Espíndola *et al.* (2018) e Medeiros (2018) utilizam o termo “dados científicos”; portanto, quando essa nomenclatura aparecer no texto desta dissertação, será em citações diretas desses autores. Os dados científicos, no entendimento do autor deste trabalho, restringiriam os tipos de dados a serem depositados nos repositórios.

## 2 FUNDAMENTAÇÃO TEÓRICA

A fundamentação teórica aborda os temas importantes e necessários para a concepção de um serviço de repositório de dados, ajuda na compreensão e no entendimento do problema e dos objetivos. As seções abaixo vão em ordem decrescente, desde a atividade científica, o contexto e o conceito de *e-Science*, os dados de pesquisa, o ciclo de vida dos dados, a gestão de dados de pesquisa, os princípios FAIR, as licenças abertas, até a certificação CoreTrustSeal em busca de repositórios digitais confiáveis, que são temas fundamentais para a implantação de repositórios de dados de acordo com a qualidade recomendada na literatura científica.

### 2.1 A ATIVIDADE CIENTÍFICA

A ciência é estruturada pelos conhecimentos produzidos por milhões de cientistas e pesquisadores ao longo dos anos, a chamada comunidade científica. O conhecimento ou o conhecimento científico, como é definido, é a base de sustentação da ciência desde o surgimento das primeiras universidades na Idade Média, fortalecido com a invenção da imprensa no século XV e refortalecido com o surgimento das tecnologias após metade do século passado. Entretanto, o processo de comunicação científica permaneceu quase inalterado desde o surgimento dos primeiros periódicos científicos no século XVII, sendo ele constituído de submissão, revisão pelos pares e publicação impressa, hoje, também digital.

A ciência é esse complexo campo que admite e comporta todas as áreas do conhecimento. Para Ziman (1979, p. 27), “[...] a admissão que o conhecimento científico deve ser público e aceito pelo consenso permite-nos traçar as complexas relações internas entre suas várias facetas”. A ideia de um conhecimento público pode ser interpretada à luz das iniciativas da ciência aberta, que ganharam corpo e práticas somente duas décadas após Ziman ter publicado sua obra.

A ciência é, inegavelmente, um produto consciente da humanidade, com suas origens históricas bem documentadas, um escopo e um conteúdo bem definidos; além do mais, conta com praticantes e expoentes reconhecidamente profissionais. [...] Ciência [...] é precisa, metódica, acadêmica, lógica e prática. A própria facilidade que encontramos de entendê-la, de focalizar as coisas com nitidez, faz-nos sentir que se trata de instrumento real, sólido, concreto (ZIMAN, 1979, p. 17).

Para Ziman (1979, p 24), “[...] a ciência não significa simplesmente conhecimentos ou informações publicados, ou seja, artigos científicos, livros, ensaios e muitos outros tipos de textos científicos”. A ciência é bem maior que isso, é o que sustenta os progressos e as descobertas em todo o planeta. É na ciência que se descobre a cura de doenças, se constroem



satélites que permitem o uso de Global Positioning System (GPS), se aprimoram e desenvolvem novos materiais e também avançam no entendimento/amadurecimento de disciplinas científicas. A ciência é movida por construções e desconstruções, o que Kuhn (2018) denominou de paradigmas científicos. A ciência possibilita que a vida dos seres humanos tenha maior utilidade laborativa e que consigam ter mais expectativa de vida que os seus ancestrais.

Ao mesmo tempo que a ciência representa estrutura grandiosa, capaz de trazer benefícios a bilhões de pessoas, é “coisa humilde”, pois a verdade é inatingível. Lida com hipóteses, teorias e modelos sempre provisórios. Nada é definitivo. E é exatamente esta possibilidade inesgotável de novidades, surpresas e coisas novas que vão ser desvendadas a responsável pela grandiosidade da ciência. Mas também, responsável por sua infinitude e complexidade. A ciência recorre, inevitavelmente, à informação e à comunicação. No entanto, é insensato restringir a comunicação à mera troca de informações entre cientistas, pois a ciência como sistema social integra elementos que vão desde a figura do pesquisador/cientista/acadêmico ao fluxo de ideias, fatos, teorias, métodos, literatura científica e instrumentos que permitem a operacionalização das investigações (TARGINO, 2000, p. 24).

Ziman (1979, p. 24) ainda diz que “[...] qualquer pessoa que tenha condições financeiras pode escrever suas ideias e a colocar num papel, imprimir e distribuí-lo”. Por isso que a ciência não deve ser vulgarizada, achando que qualquer manuscrito/livro pode ser considerado científico. Para o fazer científico é necessário seguir todos os passos e protocolos ensinados com rigor e ética aos alunos iniciantes em pesquisa desde a graduação, passando pelo mestrado, doutorado e seguindo por toda a vida do pesquisador.

O conhecimento científico é mais do que isso. Seus fatos e teorias têm de passar por um crivo, por uma fase de análises críticas e de provas, realizadas por outros indivíduos competentes e desinteressados, os quais deverão determinar se eles são bastante convincentes para que possam ser universalmente aceitos. O objetivo da Ciência não é apenas adquirir informação, nem enunciar postulados indiscutíveis; sua meta é alcançar um consenso de opinião racional que abranja o mais vasto campo possível (ZIMAN, 1979, p. 24).

A junção de cientistas de uma área é chamada de sociedade científica, ela é a base para o desenvolvimento de novos conhecimentos, segundo Freire (2006, p. 8):

[...] a base da sociedade atual começou a ser construída quando as sociedades científicas foram criadas e os primeiros periódicos científicos foram publicados, dando início à formalização do processo de comunicação científica.

Para Targino (2000, p. 13),

[...] da mesma forma que todos os grupos sociais, sem exceção, tem regras implícita ou explícitas de atuação, ou seja, uma ética reguladora mais ou menos rígida, a

comunidade científica, como estrutura social que é, não pode prescindir de valores éticos e morais.

Para a autora, a comunicação científica, como parte integrante dessa estrutura, também está sujeita à interferência de prescrições que direcionam as atitudes comportamentais dos pesquisadores e, portanto, influenciam a produção científica.

A comunicação científica é o modo que os cientistas usam para se comunicar e informar aos seus colegas o que estão produzindo, estudando, desenvolvendo, modificando, entre outras inúmeras coisas que fazem dentro dos laboratórios e grupos de pesquisa. A comunicação é realizada por meio de artigos, comunicações orais em eventos, anais, relatórios, teses e dissertações. Segundo Ziman (1979, p. 15):

A pesquisa científica é uma atividade altamente consciente, deliberada e racional, representando a antítese daqueles modelos de comportamento tradicionais, semirracionalizados e quase inconscientes que constituem o objeto convencional de estudo de antropólogos e sociólogos.

A autoria do termo **comunicação científica** é imputada a John Desmond Bernal. No capítulo *Comunicação científica* do livro *A função Social da Ciência* (1939), ele incorpora as atividades associadas à produção, disseminação e uso da informação, desde o momento da concepção da ideia pelo cientista até a informação referente aos resultados alcançados ser aceita como constituinte do estoque universal de conhecimentos pelos pares. A abrangência do termo se amplia para além das fronteiras da comunidade científica quando defende que, tanto o cientista quanto o público leigo, receberiam as informações necessárias e úteis para o desenvolvimento de seus trabalhos ou de suas atividades cotidianas. Neste sentido, a informação científica fluiria, não somente entre os cientistas, mas chegaria também até o público em geral (CARIBÉ, 2015, p. 90).

A autora cita outras nomenclaturas correlacionadas e/ou subordinadas ao termo comunicação científica, como: difusão científica, divulgação científica, popularização da ciência, disseminação científica: “Esses termos são vistos como processos, ou seja, atividades desenvolvidas com o objetivo de levar a informação científica ao cliente, o grupo social” (CARIBÉ, 2015, p. 90). Portanto, a comunicação científica é o mecanismo que a ciência criou para se sustentar ao longo dos séculos, se fortalecer e permitir que as ideias científicas não se perdessem, podendo, assim, dar os devidos créditos aos seus criadores.

Conforme apontam Freitas e Leite (2018), a comunicação científica constitui um sistema, estruturado ao longo do tempo por meio da adoção de vários processos que, em razão de diversos fatores de ordem social, econômica e tecnológica, sofreram e continuam a sofrer mudanças, sendo possível colocar em perspectiva seus elementos e as relações entre eles.

A comunicação científica, inclusive a revisão por pares, também está passando por mudanças fundamentais. As bibliotecas digitais públicas estão tirando das bibliotecas convencionais a função de preservar publicações – por causa da despesa, da disponibilidade de acesso a qualquer momento e da necessidade de manter juntos os dados experimentais e os dados sobre os dados (BELL, 2011, p. 14).

Segundo Kuramoto (2015), o Movimento Open Access, ou Movimento de Acesso Livre, introduziu novas alternativas de acesso à informação científica aos pesquisadores radicados em qualquer parte do globo terrestre e inseriu novas ferramentas de disseminação da produção científica, como, os repositórios digitais e as revistas científicas de acesso livre. O acesso aberto é importante para equiparar o acesso/disponibilidade dos conteúdos produzidos pelas universidades e pelos órgãos de pesquisa, principalmente, aquelas financiadas com o dinheiro público<sup>3</sup>.

Percebe-se que a tecnologia modificou o modo com que as pessoas se relacionam com o mundo. Os computadores pessoais e *gadgets*<sup>4</sup> ampliaram o modo de acesso às informações, assim como, a maneira de estudar, de desenvolver pesquisas e de publicá-las – a comunicação científica. As palavras em alta para a comunidade científica são compartilhamento, reuso e gestão de dados, pois agora percebeu-se a necessidade da abertura e do compartilhamento também dos dados de pesquisa em sua totalidade.

O registo das observações, ensaios e experiências, ou seja, a produção de dados, é já há vários séculos uma das características essenciais da ciência moderna. A forma e o volume desses registos ou dados científicos foram naturalmente evoluindo, crescendo em dimensão e complexidade, de acordo com a própria evolução da investigação científica, dos seus objectos [sic], metodologias e instrumentos. De igual modo, foram-se registando alterações nas formas de armazenar, preservar, aceder e partilhar os dados produzidos no âmbito da actividade [sic] científica (RIBEIRO *et al.*, 2010, p. 10).

Conforme Mueller (2007), os CI que estudam a comunicação científica, estão, geralmente, interessados em entenderem como ocorre a passagem da concepção de uma ideia original pelo cientista até a produção do conhecimento científico e de sua absorção pela

---

<sup>3</sup> Portaria Capes n. 13 de 2006, que obriga/recomenda o depósito legal de teses e dissertações em formato online em repositórios institucionais. É obrigatório, principalmente, às que receberam bolsas de pesquisa, porém, o entendimento hoje é que o programa recebe verbas e os professores são pagos com o dinheiro público, logo, as pesquisas precisam estar disponíveis para a sociedade. Disponível em: [https://www.capes.gov.br/images/stories/download/legislacao/Portaria\\_013\\_2006.pdf](https://www.capes.gov.br/images/stories/download/legislacao/Portaria_013_2006.pdf). Acesso em: 30 nov. 2020.

<sup>4</sup> *Gadgets* são dispositivos eletrônicos portáteis, criados para facilitar funções específicas e úteis no cotidiano, são exemplos os *smartphones*, *smartwatches*, *e-readers* e *videogames*.

comunidade científica. Esses pesquisadores se interessam ainda em estudarem como a informação científica é divulgada para a sociedade em geral.

Para Gomes e Rosa (2017), a informação científica é o elemento principal para o desenvolvimento científico e tecnológico e está diretamente relacionada à divulgação da informação e à construção de novos conhecimentos. Segundo Marra (2012, p. 191), “[...] o processo de comunicação científica tem sido significativamente impactado pela publicação eletrônica”.

Com a percepção que é necessário a divulgação, a ampliação e o acesso dos países pobres e em desenvolvimento às publicações científicas, o MAA ganhou força e se solidificou. Ainda de acordo com Marra (2012), os investimentos públicos em Pesquisa e Desenvolvimento (P&D) são amplamente direcionados ao setor acadêmico, mais especificamente ao financiamento de pesquisas na pós-graduação, dessa forma, para a autora, a disseminação gratuita da informação científica deveria ser uma obrigação.

Em continuidade a esse processo da comunicação científica, a abertura/publicação dos dados adentrou como elemento essencial para as sociedades científicas. Os dados de pesquisa são os elementos de maior importância para o fazer científico atual. Portanto, na comunicação científica também é exigido mudanças e novas práticas para atender o acelerado desenvolvimento científico e tecnológico do mundo.

### **2.1.1 A Ciência da Informação no uso dos dados**

A área da Ciência da Informação (CI) já estuda a comunicação científica há muitos anos, suas inovações, progressos e novas práticas, assim, os repositórios de dados são uma recente inovação no fazer científico a ser estudado pela área. Segundo Monteiro (2017, p. 15), “[...] os dados científicos estão adquirindo relevância nos estudos e práticas da CI, pois, o foco vem sendo ampliado da informação para dados na geração de conhecimento”.

A ciência do século XX foi marcada pelos avanços científicos e tecnológicos, principalmente dos países ricos durante a Segunda Guerra Mundial, com imensa produção de informações, conhecimentos e tecnologias para obterem hegemonia e vencerem a guerra, surgindo sistemas de recuperação da informação, sendo um dos fatos que contribuíram para o surgimento e consolidação da CI.

Para Robredo (2003), o fenômeno ligado ao crescimento exponencial da produção e uso de documentos, em especial nas áreas técnico-científicas, envolvia uma diversidade considerável de profissionais, em diferentes contextos, assim, a Ciência da Informação surgiu

a partir da necessidade de representar, disseminar e recuperar a crescente produção científica, facilitando o acesso aos seus conteúdos de forma efetiva.

Borko (1968) argumenta que a CI é uma ciência que investiga as propriedades e o comportamento da informação, as forças que governam o fluxo, o uso da informação e as técnicas, tanto manuais quanto mecânicas de processamento dessas para melhor armazenamento, recuperação e disseminação.

Ciência da Informação é a disciplina que investiga as propriedades e o comportamento informacional, as forças que governam os fluxos de informação, e os significados do processamento da informação, visando à acessibilidade e a usabilidade ótima. A Ciência da Informação está preocupada com o corpo de conhecimentos relacionados à origem, coleção, organização, armazenamento, recuperação, interpretação, transmissão, transformação, e utilização da informação. Isto inclui a pesquisa sobre a representação da informação em ambos os sistemas, tanto naturais quanto artificiais, o uso de códigos para a transmissão eficiente da mensagem, bem como o estudo do processamento e de técnicas aplicadas aos computadores e seus sistemas de programação. É uma ciência interdisciplinar derivada de campos relacionados, tais como a Matemática, Lógica, Linguística, Psicologia, Ciência da Computação, Engenharia da Produção, Artes Gráficas, Comunicação, Biblioteconomia, Administração, e outros campos científicos semelhantes. Têm ambos os componentes, de ciência pura visto que investiga seu objeto sem considerar sua aplicação, e um componente de ciência aplicada, visto que desenvolve serviços e produtos (BORKO, 1968, p. 1-2).

Buckland (1991) aborda três diferentes usos para a informação dentro da CI. A informação-como-processo, é quando a informação é usada para o ato de informar, assim, o conhecimento inicial do indivíduo é modificado. A informação-como-conhecimento, segundo Buckland (1991, p. 1), “[...] é aquilo que é usado para denotar a informação-como-processo”, ou seja, é o que é transmitido ao indivíduo que o faz obter conhecimento. E por fim, a informação-como-coisa que, segundo o autor, “[...] é de interesse especial no estudo de sistemas de informação”, pois, como ele completa dizendo: “[...] qualquer expressão, descrição ou representação seria ‘informação-como-coisa’” (BUCKLAND, 1993, p. 3).

Muitas definições têm sido propostas para “informação”. Um importante uso de informação é denotar conhecimento comunicado; outro é denotar o processo de informação. [...] Informação-como-coisa merece uma análise cuidadosa, parcialmente porque essa é a única forma de informação que é diretamente tratada pelos sistemas de informação. Pessoas são informadas não somente através de comunicados intencionais, mas devido a uma extensa variedade de objetos e eventos. Ser “informativo” é circunstancial e seria imprudente afirmar que alguma coisa não pode ser informativa, e consequentemente informação, em uma situação concebível. Variedades de “informação-como-coisa” diferem em suas características físicas e assim não são igualmente processadas para armazenamento e recuperação. Há considerável chance em substituí-la por representações (BUCKLAND, 1991, p. 14).

Percebe-se que essa definição de informação do autor é importante para esta pesquisa, por ajudar a definir os objetos que seriam informativos. Segundo Buckland (1991, p. 1, grifo

nosso), “[...] variedades de ‘informação-como-coisa’ incluem **dados**, textos, documentos, objetos e eventos”. Para que não se banalize a compreensão de Buckland (1991, p. 10), “[...] podemos dizer que alguns objetos ou documentos que pouco a pouco combinam algumas circunstâncias, situações, poderiam ser informativas, poderiam ser informação, isto é, informação-como-coisa”. Percebe-se que objetos carregam consigo significados, os objetos e arquivos digitais, analisados juntando variáveis e similaridades, também conseguem ser úteis, como no caso, os dados de pesquisa.

Para Saracevic (1995; 1996), a CI está inexoravelmente ligada à tecnologia da informação. Uma das principais subáreas de estudo da CI é a comunicação científica, e o seu desenvolvimento nas últimas décadas está ligado às tecnologias informacionais e ao avanço de produtos e serviços tecnológicos, por exemplo, os repositórios de dados, objeto deste estudo. Segundo Saracevic (1996, p. 42), três características constituem a razão de existência da área:

Três são as características gerais que constituem a razão da existência e da evolução da CI; outros campos compartilham-nas. Primeira, a CI é, por natureza, interdisciplinar, embora suas relações com outras disciplinas estejam mudando. A evolução interdisciplinar está longe de ser completada. Segunda, a CI está inexoravelmente ligada à tecnologia da informação. O imperativo tecnológico determina a CI, como ocorre também em outros campos. Em sentido amplo, o imperativo tecnológico está impondo a transformação da sociedade moderna *em sociedade da informação, era da informação ou sociedade pós-industrial*. Terceira, a CI é, juntamente com muitas outras disciplinas, uma participante ativa e deliberada na evolução da sociedade da informação. A CI teve e tem um importante papel a desempenhar por sua forte dimensão social e humana, que ultrapassa a tecnologia. Essas três características ou razões constituem o modelo para compreensão do passado, presente e futuro da CI e dos problemas e questões que ela enfrenta.

De acordo com Saracevic (1995), a CI estava alcançando um momento crítico em sua evolução, com vários tipos de pressões forçando o reexame de problemas envolvidos e as formas e meios de resolvê-los. Para o autor, as pressões eram: (1) a evolução da sociedade da informação está se acelerando no mundo desenvolvido, assim, os papéis sociais e econômicos das atividades de informação estão tornando-se mais evidentes. Cresce também a importância da área e traz novos participantes, aumentando a competição em busca de conquistar terreno; (2) o imperativo tecnológico está provendo e forçando o desenvolvimento e a aplicação de tecnologias a uma variedade sempre crescente de serviços de informação, de produtos, de sistemas e de redes; (3) as relações interdisciplinares que envolvem, de alguma maneira, todos os campos que lidam com problemas informacionais estão se alterando. Essas pressões, segundo Saracevic, trazem mais cooperação interdisciplinar, afinal, as relações estão se estreitando, o que favorece outros estudos no campo, como as bibliotecas digitais.

A tecnologia, conforme Saracevic (1995; 1996), é um fator-chave para o desenvolvimento e a consolidação da área, visão essa que hoje se concretizou e, ainda mais, após as inovações e os desenvolvimentos do Movimento de Acesso Aberto, como: documentos em hipertexto, plataformas de periódicos online, *softwares* de repositórios, gerenciadores bibliográficos e bases de dados em linha. Para o autor, as áreas que constituem o núcleo da CI (Biblioteconomia, Ciência da Computação, Ciência Cognitiva e Comunicação) são permeadas por tecnologia, portanto, suas razões de ser e de fazer estão na tecnologia e no desenvolvimento de inovações da comunicação científica.

A informação é considerada o objeto de estudo da ciência desde a denominação da área na década de 1960. O dado é considerado o menor elemento que constituiria a informação para as pessoas e os dados, reunidos, se transformam em informação aos usuários. Agora, os dados são elevados há um novo patamar, como elemento principal para as práticas de *e-Science*.

### 2.1.2 e-Science

A *e-Science* consiste no uso de computadores que conseguem capturar e armazenar imensas quantidade de dados de todos os tipos e em todas as áreas da ciência, também chamada de ciência do uso intensivo de dados. Segundo Appel (2014), o uso do termo *e-Science* diz respeito a um movimento que prevê a concepção de uma ciência atuante no uso intensivo de dados e na colaboração por meio da utilização de plataformas de pesquisa baseadas em computação avançada. O desenvolvimento tecnológico permitiu que novas ferramentas, telescópios, computadores e *softwares* pudessem capturar/coletar dados sem precedentes para serem usados em pesquisas. A *e-Science* está sendo denominada de “o quarto paradigma da ciência”, segundo Jim Gray<sup>5</sup> (2007).

Por ora, entende-se a *e-Science* como um conceito agregado à ideia de informação, tecnologia e dados, sendo utilizada para representar coisas, ações, métodos e sistemas sob a dicotomia do hardware e do software em âmbito científico. Acredita-se que a *e-Science* esteja preocupada com a natureza e com a significância do uso das tecnologias de dados em relação às práticas e às teorias científicas contemporâneas, aplicadas pelas mais diversas áreas do conhecimento para promover a colaboração científica em escala global por meio do uso intensivo de dados (SEMELER, 2017, p. 52-53)

---

<sup>5</sup>A citação refere-se à transcrição de uma palestra ministrada por Gray em 11 de janeiro de 2007. Porém, presente na obra *O quarto paradigma: descobertas científicas na era da e-Science*, de 2011. O livro está organizado e assinado por outros pesquisadores, inclusive o capítulo da transcrição da palestra, entretanto, decidiu-se citar o autor principal da palestra e não os organizadores do livro e da transcrição.

Segundo Kuhn (2018), o primeiro paradigma científico consistiu na ciência empírica, ou seja, nas coisas que o homem observava e descrevia, os fatos e os fenômenos que aconteciam ao seu redor. O segundo paradigma é a ciência teórica, que é a busca de uma teoria explicativa ou um modelo para fatos e fenômenos, visa fundamentar e explicar por meio racional. O terceiro paradigma científico é o uso da ciência computacional, que utiliza de instrumentos tecnológicos e computadores para encontrar as respostas científicas e estudar o mundo.

Da mesma forma que a ciência tradicional, uma ciência intensiva em dados, continua utilizando os meios formais de publicação acadêmica. Contudo, adiciona um novo elemento: a necessidade de novas formas de disseminação dos resultados de pesquisa, apresentando assim, novo desafio. De um lado, para os pesquisadores que estão preocupados com a realização das pesquisas e não com as formas de disponibilização e compartilhamento dos dados, mas que agora estão pressionados a disponibilizar seus dados de pesquisa com vistas à reutilização (ALMEIDA, 2019, p. 18).

De acordo com Semeler (2017), a *e-Science* está associada a estudos realizados em ambientes de rede, utiliza-se de ferramentas como a internet para envolver pesquisadores em ambientes de colaboração global. Segundo Oliveira e Silva (2016, p. 5),

[...] as práticas da *e-Science* como uso e reuso dos dados científicos se constituem um novo fazer científico que conduz à reflexão sobre novos arcabouços normativos, legais, institucionais e tecnológicos voltados para uma ciência aberta.

A Ciência Aberta representa, assim, uma mudança na forma de desenvolver pesquisas, com forte ênfase na disseminação do conhecimento, usando as facilidades de comunicação trazidas pelas tecnologias digitais. Para além das práticas usuais de publicar resultados da pesquisa em artigos científicos, busca-se também, e ativamente, o compartilhamento e uso de todo o conhecimento disponível desde os estágios iniciais de uma pesquisa (FAPESP, [2019], não paginado).

Dessa forma, são exigidas novas demandas da comunicação científica, muito ligada ao uso de tecnologias e de instrumentos sofisticados para a construção de conhecimentos. Sobre a organização e preservação dos dados:

[...] a ciência necessita que os dados sejam organizados e preservados durante a fase de pesquisa e compartilhados depois do fim do projeto. Mas o surpreendente não é somente a quantidade de dados, e sim tudo o que podemos realizar com eles, pois os novos avanços em mineração e visualização nos proporcionam novas formas de extrair informação útil a partir de conjuntos cada vez maiores de dados (SILVA, 2019, p. 2).

Para Oliveira e Silva (2016), a ciência aberta é o fio condutor de investigações científicas apoiadas por uma ciberinfraestrutura tecnológica e metodológica que permite o uso,



reuso e reprodutibilidade de dados de pesquisa. Segundo Henning *et al.* (2019), as novas formas de fazer ciência, pautadas fortemente no compartilhamento e no reuso de dados de pesquisa, vêm colocando em evidência a necessidade de deixar para trás a ideia dos dados apenas como insumos intermediários das atividades científicas. Os dados sempre foram importantes, mas devido à capacidade de obtenção de imensos volumes em armazená-los e a sofisticação de *softwares* de análise desses dados, obtendo infinitas possibilidades de novas contextualizações, esses ganham uma nova importância para a ciência.

No campo acadêmico, o intercâmbio de dados é de vital importância para a ciência, pois permite que os pesquisadores utilizem, analisem e reproduzam os dados de outros para replicá-los ou refutá-los com o conhecimento da causa, já que é possível ter acesso aos dados que sustentam uma investigação (SILVA, 2019, p. VI).

Percebe-se que a *e-Science* é caracterizada pelo intensivo uso de instrumentos de coleta de dados, por exemplo: sensores, telescópios, satélites, sondas espaciais, aparelhos de ressonância e demais aparelhos de medicina nuclear, ou seja, instrumentos que coletam diariamente grandes volumes de dados, em todas as áreas do conhecimento científico, são dados ambientais, climáticos, espaciais, médicos, entre outros inúmeros tipos (HEY; TANSLEY; TOLLE, 2011).

De acordo com Oliveira e Silva (2016), o quarto paradigma surge como uma nova perspectiva para a ciência aberta. Tem em seu arcabouço um volume intensivo de dados de pesquisa produzidos em grande escala, apoiado pelas Tecnologias de Informação e Comunicação (TICs) e ferramentas abertas de alta performance, além de apresentar características aprimoradas da ciência tradicional, tais como: compartilhamento, colaboração, preservação e reuso dos dados de pesquisa.

## 2.2 DADOS DE PESQUISA

Os dados de pesquisa sempre existiram e subsidiaram os estudos, porém, só após serem contextualizados e tratados que se transformam em informação, e durante o século passado essa foi considerada mais importante que os dados brutos. Com o surgimento da *e-Science*, a grande produção de dados agora possui um potencial infinito de novas contribuições e inovações científicas. Os dados são os elementos essenciais que devem ser armazenados, compartilhados e preservados. O dado é o menor elemento que, após analisado, transforma-se em informação e, na sequência, em conhecimento. Segundo Almeida (2019, p. 24), “[...] apesar de os dados estarem presentes em todos os lugares, o termo pode ter significados diferentes dependendo do

contexto, sendo usado, às vezes, como sinônimo de informação, outras se referindo apenas a números”.

Dados de pesquisa são os materiais comumente registrados e aceitos na comunidade científica como necessários para validar os resultados de pesquisa e incluem: fatos e estatísticas recolhidas para posterior referência ou análise, documentos (texto, Word), planilhas (Excel, etc.), cadernos de laboratório, cadernos de campo, diários, questionários, transcrições, fitas de áudio, fitas de vídeo, fotografias, filmes, sequências de proteínas ou genéticos, respostas de teste, slides, artefatos, amostras, coleção de objetos digitais adquiridos e gerados durante o processo de pesquisa, conteúdos de banco de dados (vídeo, áudio, texto, imagens), modelos, algoritmos, scripts, arquivos de log, software de simulação, metodologias e fluxos de trabalho, procedimentos operacionais, padrões e protocolos (DUDZIAK, 2016, não paginado).

Os dados de pesquisa podem ser coletados já em formato digital, utilizando-se de tecnologias, como *softwares* e demais instrumentos, mas também de forma manual (depois eles serão tratados e transformados para o formato digital), por inúmeros cientistas, que irão armazená-las em um único *database*, por isso, é uma prática cooperativa e colaborativa. Para Silva (2019, p. 7), “[...] os dados científicos são uma parte essencial das provas necessárias para avaliar os resultados de pesquisa e para reconstruir os fatos e processos que conduzam a tais resultados”.

Para que sejam úteis a comunidade científica, porém, os dados devem seguir uma estrutura e organização clara, e constituir coleções informativas relacionadas e registradas em formato adequado ao tema tratado, isto é, no contexto de uma determinada comunicação científica. Dessa forma, dos resultados gerados em uma pesquisa se obterá um conjunto de dados que poderá ser armazenado e reutilizado ao distribuir-se a outros pesquisadores, e inclusive poderá ampliar-se às áreas distantes dos objetivos iniciais da pesquisa (SILVA, 2019, p. 5).

Segundo Monteiro (2017), os dados estão adquirindo relevância nas pesquisas e práticas da CI, com foco no estudo dos dados para a geração de conhecimento, assim, estes preocupam-se com o armazenamento, a semântica, a preservação e a descrição dos metadados. Definir com precisão o que são dados de pesquisa é um desafio, pois eles são extremamente diferentes para cada área de pesquisa, variando o formato, o modo de coleta, o volume e a quantidade.

Hoje, há um tsunami de dados ameaçando assoberbar-nos, consumindo nossa atenção pelo próprio volume e diversidade. Alimentado por sensores aparentemente ubíquos, redes de banda larga e sistemas de armazenamento de alta capacidade, o tsunami incorpora dados de sensores que monitoram nosso Planeta nas profundezas do oceano, de instrumentos terrestres e de sistemas de imagens baseados no espaço. Inclui também medições ambientais e dados de serviços de saúde que quantificam processos biológicos e efeitos de condições ambientais. Estamos simplesmente caminhando da escassez de dados para a pletora de dados, o que leva à relativa falta da atenção

humana a qualquer dado individual e requer a assistência de máquinas para separar o joio do trigo (GANNON; REED, 2011, p. 147).

Conforme Oliveira e Silva (2016), o dado é o elemento central e o recurso principal. As mudanças nas práticas científicas têm gerado alterações no *status quo* da ciência tradicional. O novo fazer científico, denominado *e-Science*, prioriza a colaboração, o compartilhamento e o gerenciamento dos dados científicos como aspectos principais de uma ciência colaborativa.

A noção de dados pode variar consideravelmente entre pesquisadores e, ainda mais, entre áreas do conhecimento. A constatação de que **os dados são gerados para diferentes propósitos, por diferentes comunidades acadêmicas e científicas e por meio de diferentes processos** intensifica ainda mais essa percepção de diversidade (SAYÃO; SALES, 2015, p. 7, grifo do autor).

Antes da *e-Science* e da Ciência Aberta, os dados ficavam restritos aos pesquisadores que os coletavam. Conheciam-se somente aqueles dados utilizados em publicações formais (artigos, livros, teses, dissertações, entre outros). Os dados coletados são infinitamente maiores que os utilizados nas publicações, ao longo dos anos percebeu-se que outros pesquisadores tinham de recoletar (às vezes, os mesmos) dados para novas pesquisas, aumentando os custos e dificultando o desenvolvimento/progresso da ciência.

Sem dúvida, a abertura dos dados científicos representa enormes benefícios para a humanidade. No plano político, há forte consenso no sentido de que o direito de acesso à informação seja considerado um direito humano fundamental. A essa concepção se soma o aumento das iniciativas *open* na ciência, que tem levado ao surgimento de novos paradigmas para produção e distribuição do conhecimento (MACHADO, 2015, p. 207).

Pavão *et al.* (2019, p. 7) definem os dados de pesquisa como “[...] dados coletados, observados ou produzidos durante a pesquisa (números, textos, imagem, som, saídas de equipamentos) para fins de análise e produção de resultados de pesquisa originais”. Existe uma infinidade de tipos de dados de pesquisa, porque são elementos presentes em todas as áreas acadêmicas.

Podem ser de tipos numérico, descritivo ou visual e reproduzir-se em formato de papel (incluindo notas de pesquisa em cadernos, fotografias, etc.) ou digital. O conceito de dados científicos também faz referência às distintas ferramentas como protocolos, códigos numéricos, gráficos e tabelas que são necessárias para recolher e organizar os dados, tanto em trabalhos de campo quanto em laboratório. Incluem não somente os materiais e amostras biológicas e/ou ambientais extraídas, mas também os resumos gerados durante o transcurso de realização de uma pesquisa (SILVA, 2019, p. 21).

A publicação na íntegra dos dados de pesquisa é uma exigência principalmente de agências e de órgãos de fomento, que entendem que o direito à informação é um bem/produto essencial para a inclusão de pessoas e para o desenvolvimento científico e tecnológico das nações. Para isso acontecer, os dados precisam estar íntegros, padronizados e completos. Segundo o Core Trust Seal (2022, não paginado, tradução nossa): os dados criados e usados pelos cientistas devem ser gerenciados, organizados e arquivados de forma a preservar o investimento inicial em coletá-los. Os pesquisadores devem ter certeza de que os dados mantidos em arquivos permanecem úteis e significativos no futuro. As autoridades financiadoras exigem cada vez mais acesso contínuo aos dados produzidos pelos projetos que financiam e fizeram disso um elemento importante nos Planos de Gerenciamento de Dados (PGDs).

Autores e entidades classificam os dados conforme a origem, o formato ou a finalidade para a qual foi coletado. O DMPTool<sup>6</sup> classifica-os de acordo com o tipo de metodologia adotada na sua coleta, podendo ser: observacional, experimental, simulação e derivado/compilado.

Quadro 1 – Fonte dos dados de pesquisa

<b>OBSERVACIONAL</b>	<b>EXPERIMENTAL</b>
<ul style="list-style-type: none"> <li>• Capturado em tempo real, geralmente, fora do laboratório;</li> <li>• Normalmente insubstituível e, portanto, o mais importante de proteger;</li> <li>• Exemplos: leituras de sensores, telemetria, resultados de pesquisas e imagens.</li> </ul>	<ul style="list-style-type: none"> <li>• Normalmente gerado no laboratório ou sob condições controladas;</li> <li>• Muitas vezes reproduzível, mas pode ser caro ou demorado;</li> <li>• Exemplos: sequências de genes, cromatogramas e leituras de campo magnético.</li> </ul>
<b>SIMULAÇÃO</b>	<b>DERIVADO/COMPILADO</b>
<ul style="list-style-type: none"> <li>• Máquina gerada a partir de modelos de teste;</li> <li>• Provavelmente reproduzível se o modelo e as entradas forem preservados;</li> <li>• Exemplos: modelos climáticos e modelos econômicos.</li> </ul>	<ul style="list-style-type: none"> <li>• Gerado a partir de conjuntos de dados existentes;</li> <li>• Reproduzível, mas pode ser muito caro e demorado;</li> <li>• Exemplos: mineração de texto e dados, banco de dados compilado e modelos 3D.</li> </ul>

Fonte: Adaptado de DPMTTools (2022).

Para a Ferramenta DMPTool, os dados podem ser em formato: textual (notas de campo ou laboratório e respostas da pesquisa); numérico (tabelas, contagens e medições); audiovisual (imagens, gravações de som e vídeo); modelos (código de computador); específico da disciplina (FITS em astronomia, CIF em química); e específico do instrumento (saídas do equipamento). Os dados também podem ser classificados de acordo com a sua coleta, o seu

<sup>6</sup> [https://dmptool.org/general\\_guidance](https://dmptool.org/general_guidance).

caráter (primário, secundário ou terciário) e o seu grau de estruturação (estruturados e não estruturados).

Quadro 2 – Tipologia dos dados

<b>Crítérios</b>	<b>Tipos</b>	<b>Descrição</b>
Procedimento de coleta	Observacionais	Procedentes de pesquisa científica; caráter único, pois não se pode voltar a reproduzi-los; capturados em tempo real e geralmente fora de laboratório; registros de fatos ou evidências de fenômenos.
	Computacionais	Produtos da execução de modelos computacionais, simulações ou fluxos de trabalho; são reproduzíveis se preservada a documentação, o <i>hardware</i> , o <i>software</i> , os dados de entrada e os passos intermediários.
	Experimentais	Procedentes de experimentos, procedimentos realizados em condições controladas com o fim de provar ou estabelecer hipóteses; caso seja um experimento replicável, os dados são mais fáceis de reutilizar e preservar.
Caráter	Primário	Trabalhos originais de pesquisa e/ou dados brutos sem interpretação, coletados a partir de experimentos, pesquisas, entrevistas e demais técnicas; coletados a partir do problema de pesquisa; custo de obtenção maior; geralmente utilizados nas ciências, principalmente nas experimentais.
	Secundário	Menos precisos, pois não foram coletados para responder ao problema de pesquisa; dados coletados, tabulados, ordenados e disponíveis publicamente em livros, periódicos, censos, biografias, artigos, bases de dados etc.; mais utilizados nas ciências sociais; permitem a repetição dos estudos e a criação de grandes conjuntos de dados, mais ricos e sofisticados.
	Terciário	Forma de dados derivados como recontagem, categorias e resultados de dados estatísticos; frequentemente utilizados para garantir a confidencialidade dos dados primários e secundários.
Grau de estruturação	Estruturados	Facilmente transferidos a outros sistemas devido à organização segundo um modelo definido; informação armazenada em tabelas e bases de dados relacionais seguindo uma estrutura determinada onde se definem as tabelas, os campos das tabelas e as relações entre ambos.
	Semiestruturados	Tradicionalmente inclui imagens, documentos de textos e outros objetos que fazem parte de uma base de dados; não tem um modelo de dados ou uma estrutura predefinida, não sendo possível manter em uma estrutura de base relacional; são irregulares e flexíveis, frequentemente adicionados hierarquicamente; possuem um conjunto consistente de conteúdo semântico e algum meio de classificá-los e ordená-los. Exemplos: páginas da <i>web</i> que abrigam conteúdo em html e metadados entre as etiquetas.
	Não estruturados	Não possuem um modelo definido ou estrutura identificável; cada elemento individual pode ter uma estrutura ou formato; nem todos os dados dentro de um conjunto possuem a mesma estrutura. Exemplo: mensagens do Facebook, tuítes ou <i>blogs</i> .
Nível de abertura	Dados abertos	Podem ser utilizados e distribuídos por qualquer pessoa, sem barreiras técnicas ou legais, sob os requisitos de reconhecer a autoria e compartilhar o novo produto nas mesmas condições.
Formato	Arquivos digitais	O formato e o <i>software</i> com que se criam os dados dependem da forma que os pesquisadores coletam e analisam e, geralmente, são influenciados pelas normas e pelos costumes de cada disciplina; grande quantidade de formatos disponíveis; opção mais segura para garantir o acesso; e o uso de formatos padrões.

Fonte: Almeida (2019, p. 27), baseado em Corrêa (2017).

Assim, os dados de pesquisa são múltiplos e presentes em todas as áreas do conhecimento, com inúmeras variações de tipos e de formatos, mas o que têm em comum são o poder de modificar as descobertas e os avanços da sua área com a ciência aberta e o reuso desses dados, trata-se de dados FAIR, dados íntegros, seguros, preserváveis a longo prazo e interoperáveis também por máquinas.

### 2.2.1 Os princípios FAIR

Os princípios FAIR – *Findable* (encontrável), *Accessible* (acessível), *Interoperable* (interoperável) e *Reusable* (reutilizável) – auxiliam na qualidade dos dados e demais exigências para que estes sejam disponibilizados com a finalidade do reuso e com vida útil após o término das pesquisas. Os dois primeiros requisitos: encontrável e acessível (F e A) são realizados no nível dos metadados, já os dois últimos: interoperável e reutilizável (I e R) relacionam-se diretamente com os conjuntos de dados. Esses princípios são uma criação da Global Open (GO) FAIR, que a configura como “[...] uma iniciativa de baixo para cima, orientada para as partes interessadas e autogovernada que visa implementar os princípios de dados FAIR” (GO FAIR, 2022). São princípios que orientam as ações e os procedimentos que asseguram a qualidade no tratamento dos dados.

Os princípios enfatizam a capacidade de ação da máquina (ou seja, a capacidade dos sistemas computacionais de encontrar, acessar, interoperar e reutilizar dados com nenhuma ou mínima intervenção humana) porque os humanos dependem cada vez mais de suporte computacional para lidar com dados como resultado do aumento no volume, complexidade e velocidade de criação de dados (GOFAIR.org, 2022, tradução nossa).

Segundo Henning *et al.* (2019), em 2017, a Comissão Europeia passou a exigir que os dados de pesquisa depositados em repositórios atendam aos princípios FAIR, pois, anteriormente, eles não eram encontráveis, acessíveis, interoperáveis e reutilizáveis.

São dados que atendem aos “princípios FAIR” um acrônimo para *Findable* [encontrável], *Accessible* [acessível], *Interoperable* [interoperável] e *Reusable* [reutilizável], estão presentes nas discussões e práticas contemporâneas da ciência de dados, desde o início de 2014 e tiveram sua aplicação consolidada em 2017, quando a Comissão Europeia passou a exigir a adoção de plano de gestão de dados, com base nesses princípios, por projetos financiados por seus recursos. Desde então, tais princípios passaram a ser norteadores da descoberta, do acesso, da interoperabilidade, do compartilhamento e da reutilização dos dados de pesquisa (HENNING *et al.*, 2019, p. 175).

A GO FAIR possui escritórios regionais, sendo eles: GO FAIR Áustria, GO FAIR Brasil, GO FAIR Dinamarca, GO FAIR França e GO FAIR Alemanha. O escritório brasileiro, criado em 2018 e hospedado no IBICT, tem “a responsabilidade de difundir, apoiar e coordenar as atividades relacionadas à adoção da estratégia de implementação dos princípios FAIR definida pela iniciativa GO FAIR em todo o território brasileiro” (GO FAIR Brasil, 2022). Se compromete, desse modo, com as seguintes atividades no contexto da ciência e da inovação brasileira:

- Apoiar e coordenar os institutos participantes nas Redes de Implementação de acordo com seus objetivos específicos na adoção das estratégias/diretrizes de implementação aprovadas definidas pela iniciativa GO FAIR;
- Apoiar e coordenar as Redes de Implementação que estejam dispostas a definir estratégias e diretrizes para a implementação dos princípios FAIR, nos casos em que estas estratégias/diretrizes estejam ausentes;
- Revisar e consolidar as estratégias/diretrizes de implementação existentes e as criadas pelas Redes de Implementação apoiadas, garantindo que elas sigam os princípios FAIR;
- Desenvolver mecanismos de divulgação para as estratégias/diretrizes de implementação definidas pelas Redes de Implementação e aprovadas pela iniciativa GO FAIR;
- Manter constante comunicação com o Escritório Internacional de Apoio e Coordenação GO FAIR para:
  - informar sobre as atividades coordenadas pelo GO FAIR Brasil;
  - informar sobre as atividades coordenadas diretamente pelo escritório internacional e pelos outros escritórios regionais espalhados pelo mundo, e;
  - discutir com o escritório internacional como alinhar e otimizar essas diversas atividades (DECLARAÇÃO GO FAIR BRASIL, 2019, não paginado).

Os princípios FAIR definem as características que os recursos, as ferramentas, os vocabulários e a infraestrutura de dados devem exibir para contribuir com a descoberta e a reutilização de dados e metadados (MONTEIRO; SANT’ANA, 2019). Segue abaixo um resumo de como utilizar tais princípios, incluindo as ações e a maneira de tratar os dados de pesquisa, para que estejam sempre aptos à reutilização e ao reuso.

Quadro 3 – Guia dos princípios FAIR

	<b>PRINCÍPIO</b>	<b>COMO APLICAR</b>
<b>F</b>	<b>FINDABLE (ENCONTRÁVEL)</b>	<b>Para serem encontrados</b>
F1.	Os (meta)dados devem ter identificadores globais, persistentes e identificáveis.	Metadados com identificador único – Unique and persistent identifier (PID) ou equivalente, para cada conjunto de dados. Ex: DOI, ARK, RRID, PID.
F2.	Os dados devem ser descritos com metadados enriquecidos (impacta diretamente R1).	Conjuntos de dados devem possuir metadados ricos o suficiente para que, uma vez indexados para um mecanismo de busca, eles possam ajudar o usuário dos dados a encontrá-los, mesmo que esse não possua o seu identificador.

F3.	Os metadados devem incluir clara e explicitamente os identificadores dos dados que descrevem.	Como não podemos prever que os dados e seus metadados estejam sempre juntos (por exemplo, quando os metadados são indexados por um mecanismo de busca e, portanto, estão em uma plataforma diferente dos dados), a associação entre eles deve ser feita pela inclusão do identificador dos dados pelos metadados.
F4.	Os (meta)dados devem ser registrados ou indexados em recursos que ofereçam capacidades de busca.	Para que os dados sejam encontrados, seus metadados devem ser indexados por mecanismos de busca que, por sua vez, permitem aos usuários encontrá-los por meio de elementos desses metadados.
<b>A</b>	<b>ACCESSIBLE (ACESSÍVEL)</b>	<b>Para serem acessíveis</b>
A1.	Os (meta)dados devem ser recuperáveis pelos seus identificadores usando protocolo de comunicação padronizado.	Com o identificador do conjunto de dados ou de seus metadados, o usuário deve poder recuperá-los por meio de um protocolo de comunicação padronizado.
A1.1	O protocolo deve ser aberto, gratuito e universalmente implementável.	Independente de licenciamento dos dados e dos metadados, o protocolo de comunicação usado para dar acesso a eles deve ser aberto, gratuito e passível de ser implementado por qualquer interessado.
A1.2	O protocolo deve permitir procedimentos de autenticação e de autorização, quando necessário.	Quando necessário, dependendo das restrições ao acesso aos dados e/ou metadados, um mecanismo para autenticação e autorização para o acesso deve ser permitido pelo protocolo de comunicação.
A2.	Metadados devem ser acessíveis, mesmo quando os dados não estão mais disponíveis.	É preciso existir um conjunto de estratégias de preservação para os dados e metadados. Minimamente, os metadados devem ser sempre acessíveis, possibilitando a criação de índices para conjuntos de dados atuais e passados.
<b>I</b>	<b>INTEROPERABLE (INTEROPERÁVEL)</b>	<b>Para serem interoperáveis</b>
I1.	(Meta)dados devem ser representados por meio de uma linguagem formal, acessível, compartilhada e amplamente aplicável para a representação do conhecimento.	Para podermos representar dados e metadados, devemos utilizar linguagens de representação do conhecimento que sejam formais, acessíveis e amplamente aplicáveis. Por exemplo: RDF, XML, DICOM etc.
I2.	(Meta)dados devem usar vocabulários que seguem os princípios FAIR.	Como os dados e metadados devem possuir referências aos vocabulários que contêm os conceitos utilizados, devemos garantir que sejam utilizados vocabulários que também sigam os princípios FAIR.
I3.	(Meta)dados devem incluir referências qualificadas para outros (meta)dados.	Referenciar os conjuntos de dados devidamente, possibilitando que conjuntos de dados gerados, a partir de outros conjuntos de dados, sejam ligados. Possibilitar a ligação semântica entre eles por meio dos padrões adotados.
<b>R</b>	<b>REUSABLE (REUSÁVEL)</b>	<b>Para serem reutilizáveis</b>
R1.	(Meta)dados são descritos com uma pluralidade de atributos precisos e relevantes.	Prover metadados que permitam aos potenciais usuários avaliarem se é possível reusar os dados ou metadados e, também, se eles se adequam às suas necessidades.
R1.1	(Meta)dados devem ser disponibilizados com licenças de uso claras e acessíveis.	É fundamental que o responsável pelos dados e metadados defina explicitamente quem pode ter acesso a eles, para que e sob quais condições. Essas informações são definidas por meio de suas licenças de uso.



R1.2	(Meta)dados devem estar associados à sua proveniência.	Especificar a proveniência (linhagem) dos dados é importante não só para que o usuário possa avaliar a utilidade dos dados ou metadados, mas também para que possamos atribuir o devido crédito a quem produziu, manteve ou editou esses dados. Dentre as informações relevantes à proveniência destaca-se: (a) a linhagem dos dados, ou seja, o processo de obtenção do dado (gerado ou coletado); (b) particularidades ou limitações sobre os dados que outros usuários devem conhecer; (c) data da geração do conjunto de dados, condições de laboratório, quem preparou os dados, configurações de parâmetros, nome e versão do software utilizado; (d) explicitar se são dados brutos ou processados; (e) a versão dos dados arquivados e/ou reutilizados deve ser claramente especificada e documentada.
R1.3	(Meta)dados devem estar alinhados com padrões relevantes ao seu domínio.	Atender os padrões específicos da comunidade da área. Atender às boas práticas de arquivamento e ao compartilhamento do campo de pesquisa específico.

Fonte: Wilkinson *et al.* (2016 *apud* HENNING *et al.*, 2019, p. 178-179).

Pode-se sistematizar os princípios FAIR para os dados de pesquisa em repositórios da seguinte forma:

**Encontrável** – os dados precisam ser descobertos e conhecidos para serem reusados por outros pesquisadores. Esse requisito diz respeito à capacidade de localizar os dados armazenados em plataformas de repositórios. Somente com a devida descrição dos metadados, esse requisito será cumprido.

**Acessível** – está ligado a esses dados estarem depositados em acesso aberto e aptos a serem usados por outros pesquisadores. Para ser acessível, precisa-se garantir que estes sejam preservados por meio de identificares persistentes.

**Interoperável** – os dados precisam ser entendidos e reusados também por outros computadores. Atender a esse critério subentende que outro computador consiga acessar o repositório, entender os metadados e compreender todas as variáveis presentes nos arquivos em que os dados foram disponibilizados. Segundo Wilkinson *et al.* (2016, p. 1), “[...] os Princípios FAIR enfatizam especificamente o aprimoramento da capacidade das máquinas de encontrar e usar automaticamente os dados, além de apoiar sua reutilização por indivíduos”.

**Reusável** – refere-se à capacidade que os arquivos de dados e conjuntos de dados têm de serem utilizados em outras pesquisas. Por isso, a importância de salvar em formatos gerais e não proprietários, acompanhados de toda a documentação e de metadados detalhados para a compreensão das variáveis utilizadas, o contexto, o tipo de coleta, ou seja, todo o cenário da pesquisa.

## 2.3 A GESTÃO DE DADOS DE PESQUISA

A gestão de dados de pesquisa consiste em ações necessárias para que os dados tenham integridade, utilidade e qualidade. São passos que envolvem desde o planejamento da pesquisa, a execução/coleta, o formato, os *softwares* de análise e de limpeza, a descrição dos metadados, o armazenamento, até o depósito e a preservação dos dados. Segundo Silva (2019, p. 3), “[...] para os pesquisadores, uma gestão adequada dos dados de pesquisa permite novas maneiras de comparação e de descobrimentos, isto é, permite gerar novos campos de pesquisa”. Para Silva (2021, comunicação oral), “Dados são fatos. Os metadados são as descrições dos fatos [...]”.

A qualidade diz respeito tanto ao conjunto de dados a serem depositados e armazenados quanto à descrição dos metadados, para que esses dados sejam encontráveis e reutilizáveis. Segundo Sayão e Sales (2015, p. 11), “[...] os dados e as coleções de dados de pesquisa possuem um tempo de vida maior que os projetos de pesquisa que os criaram”. Dessa maneira, os pesquisadores continuam a utilizar esses dados em publicações e outras pesquisas, por isso, é importante que sejam armazenados corretamente (com a finalidade da preservação digital) e descritos com algum padrão internacional de metadados (com a finalidade da interoperabilidade).

A gestão dos dados é o planejamento de toda a vida útil dos dados, desde antes da coleta, pois, exige-se definir previamente o tipo de coleta, detalhar os instrumentos que serão utilizados, utilização de metadados padrões, de metadados internacionais e de metadados específicos para cada área do conhecimento (ver Anexo A). Convém mencionar que o tipo de coleta muda dependendo da área, por exemplo: na Meteorologia não se coleta dados da mesma forma que na Psicologia, mas devem seguir as mesmas etapas de planejamento, de coleta, de tratativa, de armazenamento e de descrição dos metadados, para que outros pesquisadores consigam compreender os dados e reutilizá-los. Essa gestão de dados pelos bibliotecários também pode ser chamada de curadoria de dados, gerando um novo campo para atuação do profissional.

Para o encaminhamento das atividades científicas, na atualidade, a gestão de dados de pesquisa em um ambiente distribuído e em rede é considerada fundamental e torna-se um desafio crescente para o mundo da pesquisa da Ciência da Informação e da Computação. Esta gestão deve promover a identificação de princípios teóricos que apoiem os processos de organização, significação e recuperação, dando coerência, com o desenvolvimento de normas e métricas de apoio. Tudo isto apoiado por plataformas tecnológicas que impulsionem a curadoria dos dados e viabilizem o compartilhamento de dados de forma eficaz, para que esses recursos atinjam todo seu potencial de reuso (HENNING *et al.*, 2019, p. 176).

Percebe-se que são ações/passos essenciais com o propósito do compartilhamento e do reuso. Neste caso, disponibilizar dados íntegros aos demais colegas/pesquisadores, vai muito além de somente coletar e depois depositar em um repositório, a interoperabilidade permite que os dados sejam migrados em caso de obsolescência ou troca de plataforma e o reuso em novas pesquisas. Deve-se buscar toda a segurança de que dados são realmente aqueles que foram coletados, que não ocorreu alguma inconsistência ou perda. Além disso, a gestão de dados busca assegurar os formatos de armazenamento, as cópias de segurança, a equipe responsável pela gestão de dados, os identificadores persistentes, a propriedade intelectual e a preservação digital a curto, médio e longo prazo.

A gestão de dados científicos é um conjunto de atividades que visa a coletar, armazenar, gerenciar e compartilhar dados provenientes de pesquisa científica. Uma gestão de dados eficaz possibilita a racionalização de recursos, por meio do reuso e compartilhamento de dados. Na USP, a gestão de dados científicos tem a finalidade de auxiliar o pesquisador em relação a: planejamento, organização e segurança; documentação e compartilhamento; preparação dos conjuntos de dados para depósito; preservação dos dados; questões relacionadas a direitos autorais; licenciamento e propriedade intelectual (UNIVERSIDADE DE SÃO PAULO, 2020, não paginado).

Os dados depositados em repositórios de dados devem utilizar de licenças abertas para serem utilizados sem barreiras legais, sendo necessário ser citados de maneira correta, ou seja, atribuir ao autor os devidos créditos nas publicações, por exemplo, colaboração ou coautoria. Para Silva (2019, p. 33), “[...] habitualmente, os organismos financiadores de pesquisa requerem que os dados produzidos no curso de um projeto estejam disponíveis em acesso livre”.

Um dos componentes da gestão é o ciclo de vida dos dados, não é o único, mas é o principal no que se refere às fases da coleta até a preservação dos dados. A gestão constitui-se de todo o trabalho de planejamento, documentação, ciclo de vida, descrição de metadados, licenças etc.

### **2.3.1 Ciclo de vida dos dados de pesquisa (CVDP)**

O ciclo de vida são as etapas a serem seguidas pelos pesquisadores para que os dados não sejam coletados de qualquer maneira. A adoção de um modelo de ciclo de vida permite a realização da pesquisa, seguindo passos e executando boas práticas na tratativa dos dados. São muitos os modelos de ciclo de vida dos dados, a escolha, portanto, deverá ser por aquele que melhor se ajusta a área de pesquisa ou aos tipos de dados a serem coletados.

A gestão de dados científicos cobre todo o chamado “ciclo de vida” dos dados, ou seja, desde a sua coleta até o armazenamento de longo prazo, passando por uma série

de processamentos de limpeza, curadoria, anotação, indexação e transformação. Grande parte da pesquisa científica de hoje exige algum tipo de análise e processamento de dados. Com isto, o planejamento da gestão dos dados utilizados e gerados em uma pesquisa passou a fazer parte integral da metodologia científica, sendo, inclusive, considerado como um dos itens necessários de boas práticas de pesquisa (MEDEIROS, 2018, não paginado).

O ciclo de vida dos dados assegura que essas informações serão: corretas, íntegras e descritas detalhadamente, o que possibilita que sejam usadas no futuro. É praticamente uma garantia de qualidade para os dados de pesquisa. Abaixo, apresenta-se os quatro modelos de ciclo de vida dos dados mais utilizados e citados na literatura de gestão de dados.

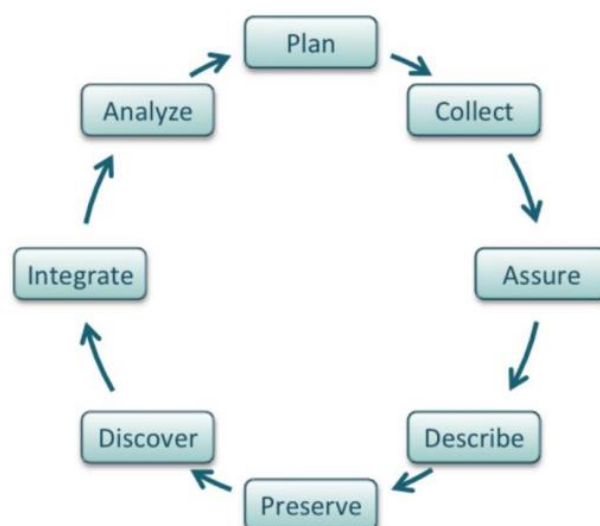
### 2.3.1.1 Ciclo de vida de Dados DATAONE

O *Data Observation Network for Earth* (DataONE) é financiado, principalmente, pela *National Science Foundation* (NSF). Segundo Strasser *et al.* (2012, p. 2),

[...] em janeiro de 2011, a *National Science Foundation* (NSF) instituiu a exigência de que um plano de gerenciamento de dados de duas páginas fosse incluído em cada proposta de subsídio como um documento suplementar.

O modelo de ciclo de vida DATAONE é cíclico e contém várias ações e práticas a serem realizadas entre as etapas.

Figura 1 – Ciclo de dados de pesquisa DATAONE



Fonte: Strasser *et al.* (2012, p. 3).

**Planejar:** descrição dos dados que serão compilados e como os dados serão gerenciados e disponibilizados ao longo de sua vida.

**Coletar:** as observações são feitas manualmente ou com sensores ou outros instrumentos e os dados são colocados em formato digital.

**Assegurar:** a qualidade dos dados é garantida por meio de verificações e inspeções.

**Descrever:** os dados são descritos com precisão e minuciosamente usando os padrões de metadados apropriados.

**Preservar:** os dados são submetidos a um arquivo de termo (ou seja, centro de dados).

**Descobrir:** os dados potencialmente úteis são localizados e obtidos, juntamente com as informações relevantes sobre os dados (metadados).

**Integrar:** dados de fontes diferentes são combinados para formar um conjunto homogêneo de dados que podem ser facilmente analisados.

**Analisar:** os dados são analisados (STRASSER *et al.*, 2012, p. 3, tradução nossa).

Algumas perguntas devem ser feitas e respondidas em cada fase do ciclo de vida dos dados e na fase do plano, segundo o modelo DATAONE (ver Apêndice A). Ao transformar o texto do modelo DATAONE em tabela (ver Apêndice A), pode-se usá-la como um *workflow* para o auxílio de pesquisadores que adotem esse ciclo de vida.

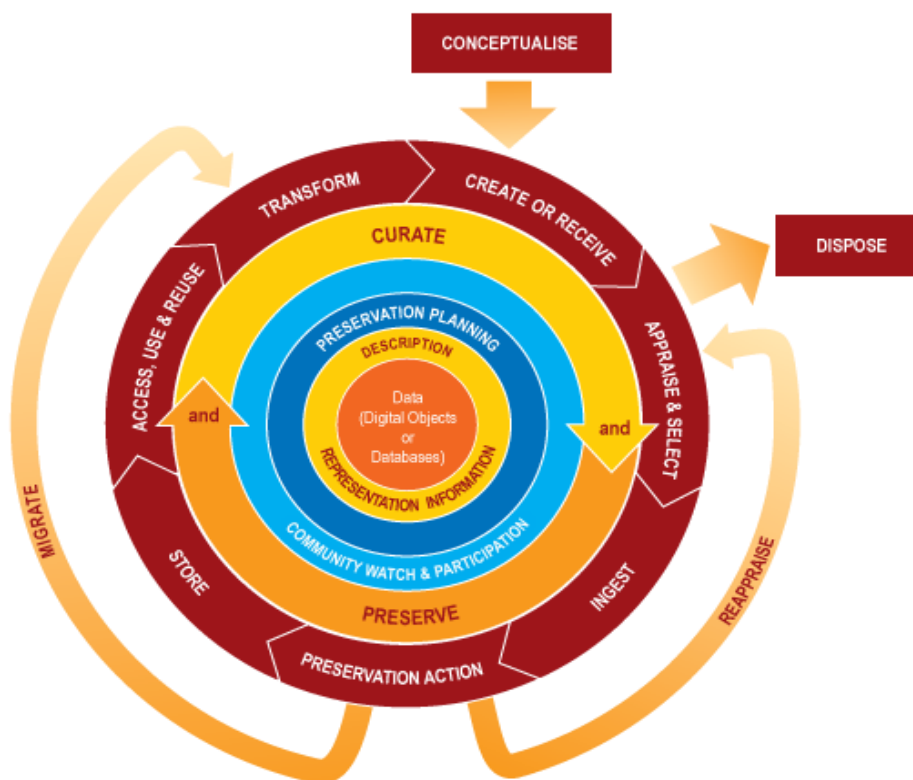
### 2.3.1.2 CURATION LIFECYCLE MODEL

O modelo *Digital Curation Centre* (DCC) é centrado nos dados e ajuda a definir fluxos de trabalho de gerenciamento de dados de pesquisa. Segundo o DCC<sup>7</sup>, o *Curation Lifecycle Model* é “uma visão geral gráfica de alto nível dos estágios necessários para o gerenciamento, curadoria e preservação de dados bem-sucedidos”.

O modelo pode ser usado para planejar atividades dentro de uma organização ou consórcio para garantir que todas as etapas necessárias sejam realizadas, cada uma na sequência correta. O modelo permite que a funcionalidade granular seja mapeada em relação a ele; para definir funções e responsabilidades e construir uma estrutura de padrões e tecnologias para implementar. Ele pode ajudar com o processo de identificar etapas adicionais que podem ser necessárias, ou ações que não são exigidas por determinadas situações ou disciplinas, e garantir que os processos e políticas sejam documentados adequadamente (CURATION LIFECYCLE MODEL, 2022, não paginado).

Para o *Curation Lifecycle Model* (2022), os dados podem ser objetos digitais e banco de dados. O modelo conceitua dados como: “[...] qualquer informação em formato digital binário, está no centro do ciclo de vida de curadoria”. E os objetos digitais, em objetos digitais simples: “[...] são itens digitais discretos; como arquivos de texto, imagem ou arquivos de som, juntamente com seus identificadores e metadados relacionados” e objetos digitais complexos: “[...] são objetos digitais discretos, feitos pela combinação de vários outros objetos digitais como sites”. O modelo divide as etapas em ações (Quadro 4).

<sup>7</sup> CURATION LIFECYCLE MODEL. Disponível em: <https://www.dcc.ac.uk/guidance/curation-lifecycle-model>. Acesso: 18 ago. 2022.

Figura 2 – *Curation Lifecycle Model*

Fonte: Curation Lifecycle Model (2022).

Quadro 4 – Etapas do *Curation Lifecycle Model*

<b>AÇÕES DO CICLO DE VIDA COMPLETO</b>	
Descrição e representação da informação	Atribuir metadados administrativos, descritivos, técnicos, estruturais e de preservação, usando padrões apropriados para garantir descrição e controle adequados a longo prazo. Coleta e atribui informações de apresentação necessárias para entender e renderizar o material digital e os metadados associados.
Planejamento da preservação	Planejar a preservação ao longo do ciclo de vida de curadoria de material digital. Isso incluiria planos para gerenciamento e administração de todas as ações do ciclo de vida de curadoria.
Observação e participação da comunidade	Ficar de olho nas atividades apropriadas da comunidade e participar do desenvolvimento de padrões compartilhados, ferramentas e <i>softwares</i> adequados.
Curar e preservar	Conhecer e realizar ações gerenciais e administrativas planejadas para promover a curadoria e a preservação ao longo do ciclo de vida da curadoria.
<b>AÇÕES SEQUENCIAIS</b>	
Conceituar	Conceber e planejar a criação de dados, incluindo método de captura e opções de armazenamento.
Criar ou receber	Criar dados, incluindo metadados administrativos, descritivos, estruturais e técnicos. Metadados de preservação também podem ser adicionados no momento da criação. Receber dados, de acordo com as políticas de coleta documentadas, de criadores de dados, outros arquivos, repositórios ou centros de dados e, se necessário, atribuir metadados apropriados.
Avalie e selecione	Avaliar os dados e selecionar para a curadoria e a preservação de longo prazo. Aderir às orientações documentadas, políticas ou requisitos legais.

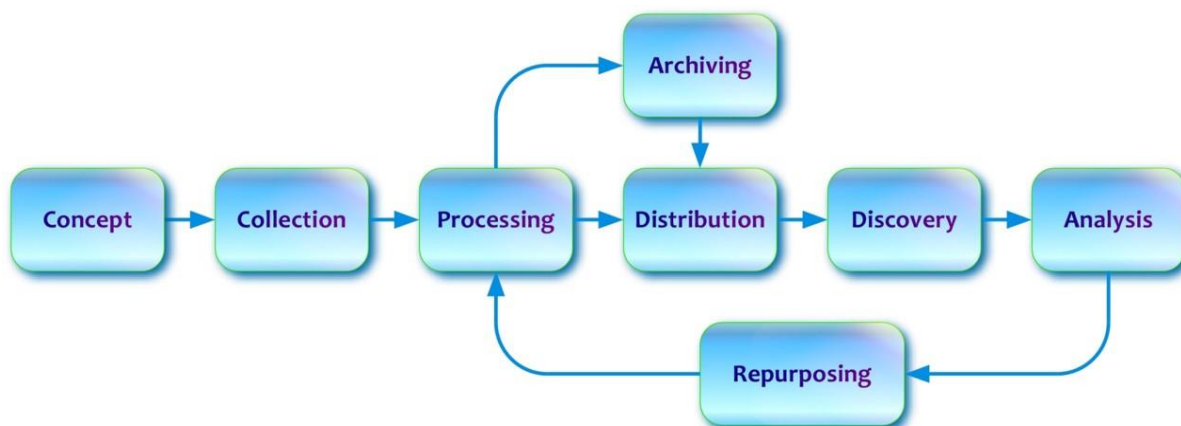
Ingerir	Transferir dados para um arquivo, repositório, centro de dados ou outro custodiante. Aderir às orientações documentadas, políticas ou requisitos legais.
Ação de preservação	Realizar ações para garantir a preservação e a retenção a longo prazo da natureza oficial dos dados. As ações de preservação devem garantir que os dados permaneçam autênticos, confiáveis e utilizáveis, mantendo sua integridade. As ações incluem limpeza de dados, validação, atribuição de metadados de preservação, atribuição de informações de representação e garantia de estruturas de dados ou formatos de arquivo aceitáveis.
Armazenar	Armazenar os dados de maneira segura, respeitando os padrões relevantes.
Acesse, use e reutilize	Garantir que os dados estejam acessíveis, tanto para usuários designados quanto para reutilizadores, diariamente. Isso pode ser na forma de informações publicamente disponíveis. Controles de acesso robustos e procedimentos de autenticação podem ser aplicáveis.
Transformar	Criar dados a partir do original, por exemplo: <ul style="list-style-type: none"> <li>- Migração para um formato diferente.</li> <li>- Criação de um subconjunto, por seleção ou consulta, para criar resultados recém-derivados, talvez para publicação.</li> </ul>
<b>AÇÕES OCASIONAIS</b>	
Descarte	Descartar os dados que não foram selecionados para curadoria e preservação de longo prazo de acordo com políticas documentadas, orientações ou requisitos legais. Normalmente, os dados podem ser transferidos para outro arquivo, repositório, centro de dados ou outro custodiante. Em alguns casos, os dados são destruídos. A natureza dos dados pode, por razões legais, exigir uma destruição segura.
Reavaliar	Devolver os dados que falham nos procedimentos de validação para posterior avaliação e resseleção.
Migrar	Migrar dados para um formato diferente. Isso pode ser feito de acordo com o ambiente de armazenamento ou para garantir a imunidade dos dados à obsolescência de <i>hardware</i> ou <i>software</i> .

Fonte: Adaptado de DCC (2022).

### 2.3.1.3 DDI LIFECYCLE MODEL

O DDI *Lifecycle* é um modelo desenvolvido pela *Data Documentation Initiative Alliance (DDI Alliance)*, sendo uma colaboração internacional que busca “estabelecer padrões de metadados e produtos semânticos para descrever dados de ciências sociais, dados que cobrem a atividade humana e outros dados baseados em métodos observacionais” (DDI, 2022). O *DDI-LifeCycle 3.3 Schema*, publicado em abril de 2020, é um *software* livre que pode ser redistribuído e/ou modificado sob a licença *Creative Commons 4.0 Internacional*, portanto, todos os seus subprodutos e modificações devem possuir a mesma licença.

Figura 3 – DDI Lifecycle

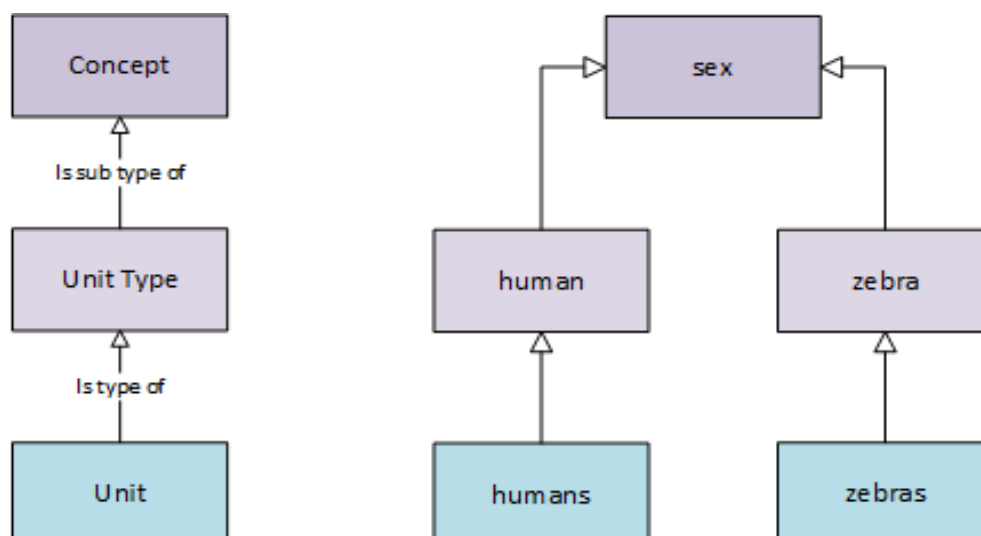


Fonte: Documentação DDI 3.3 (2020).

Os elementos que representam o modelo DDI são: o conceito (*concept*), a coleção (*collection*), o processamento (*processing*), o arquivamento (*archiving*), a distribuição (*distribution*), a descoberta (*discovery*), a análise (*analysis*) e o reaproveitamento (*repurposing*). É um modelo utilizado nas ciências humanas e sociais, pois consegue representar conceitos ligados a teorias, a fatos, a datas, a hierarquias e demais características que os dados nessas áreas possuem. Sobre a cobertura do *DDI Lifecycle* (2022, não paginado, tradução nossa):

Conceitos são uma unidade de conhecimento criada por uma combinação única de características; Tipo de Unidade é sinônimo de Objeto, que é uma classe ou objeto de interesse; a combinação permite que um único conceito seja usado entre diferentes Tipos de Unidades e, portanto, Unidades.

Figura 4 – Conceito, tipo de unidade e unidade



Fonte: Documentação DDI 3.3 (2020).



Dessa maneira, é um ciclo de vida específico para as ciências humanas e sociais, pois permite descrever relacionamentos entre pessoas e organizações e contextualizá-los no espaço e no tempo.

#### 2.3.1.4 *Research Data Lifecycle*

O *Research Data Lifecycle* é um modelo do *UK Data Service*, o principal parceiro dele é o *UK Data Archive*, fundado em 1967, na Universidade de Essex, sendo o curador da maior coleção de dados digitais de ciências sociais no Reino Unido. Esse plano inclui as seguintes etapas: planejamento de pesquisa, coleta de dados, processamento e análise de dados, publicação e compartilhamento de dados, preservação de dados e reutilização de dados.

Figura 5 – *Data Lifecycle Model*



Fonte: UK Data Service ([2022]).

Segundo o UK Data Service (2019), as ações que compreendem as etapas sugeridas são:

Quadro 5 – Etapas do *Research DATA LIFECYCLE*

Pesquisa de planejamento	Coletando dados
<ul style="list-style-type: none"> <li>• Desenhar o projeto de pesquisa;</li> <li>• Planejar o gerenciamento de dados;</li> <li>• Consentir o plano para compartilhamento;</li> <li>• Planejar a coleta de dados, protocolos de processamento e modelos;</li> <li>• Explorar fontes de dados existentes.</li> </ul>	<ul style="list-style-type: none"> <li>• Coletar dados;</li> <li>• Capturar dados com metadados;</li> <li>• Adquirir dados existentes de terceiros.</li> </ul>

<p style="text-align: center;"><b>Processamento e análise de dados</b></p> <ul style="list-style-type: none"> <li>• Inserir, digitalizar, transcrever e traduzir dados;</li> <li>• Verificar, validar, limpar e anonimizar dados;</li> <li>• Derivar dados;</li> <li>• Descrever e documentar dados;</li> <li>• Gerenciar e armazenar dados;</li> <li>• Produzir resultados de pesquisa;</li> <li>• Citar fontes de dados.</li> </ul>	<p style="text-align: center;"><b>Publicação e compartilhamento de dados</b></p> <ul style="list-style-type: none"> <li>• Estabelecer direitos autorais;</li> <li>• Criar documentação do usuário;</li> <li>• Criar metadados de descoberta;</li> <li>• Selecionar o acesso apropriado aos dados;</li> <li>• Publicar/compartilhar dados;</li> <li>• Promover dados.</li> </ul>
<p style="text-align: center;"><b>Preservando dados</b></p> <ul style="list-style-type: none"> <li>• Migrar dados para o melhor formato ou mídia;</li> <li>• Armazenar e fazer <i>backup</i> de dados;</li> <li>• Criar documentação de preservação;</li> <li>• Preservar e curar dados.</li> </ul>	<p style="text-align: center;"><b>Reutilizando dados</b></p> <ul style="list-style-type: none"> <li>• Realizar análises secundárias;</li> <li>• Realizar pesquisas de acompanhamento;</li> <li>• Fazer revisões de pesquisas;</li> <li>• Analisar as descobertas;</li> <li>• Usar dados para ensinar e aprender.</li> </ul>

Fonte: Adaptado de UK Data Service (2019)<sup>8</sup>.

Para finalizar essa primeira parte sobre a gestão de dados e os modelos de ciclos de vida de dados que podem ser adotados nas instituições e/ou pelos pesquisadores de acordo com a área e o caráter da sua pesquisa, Almeida (2019) sintetizou as atividades realizadas nos modelos de ciclo de vida de dados de pesquisa, elaborando um quadro das etapas e das atividades realizadas.

Quadro 6 – Atividades típicas realizadas no ciclo de vida dos dados de pesquisa

<b>Etapas</b>	<b>Principais atividades</b>
Descoberta e planejamento	Desenhar a pesquisa;
	Planejar a gestão de dados;
	Planejar a permissão para compartilhamento;
	Planejar a coleta de dados e os protocolos e modelos de processamento;
	Encontrar e descobrir fontes de dados existentes.
Coleta	Coletar dados-gravação, observação, medição, experimentação e simulação;
	Capturar e criar metadados;
	Adquirir dados de terceiros.
Processamento e análise	Inserir, digitalizar, transcrever e traduzir dados;
	Verificar, validar, limpar e anonimizar os dados, quando necessário;
	Derivar dados;
	Descrever e documentar;
Publicação e compartilhamento	Analisar.
	Interpretar;
	Produzir resultados de pesquisa ( <i>research outputs</i> );
	Estabelecer a autoria ( <i>authoring publications</i> );
	Citar fontes de dados;
	Gerenciar e administrar dados;
	Estabelecer os direitos autorais ( <i>copyright of data</i> );
	Criar metadados de descoberta e documentação do usuário;
	Publicar ou compartilhar dados;
	Distribuir dados;
Controlar o acesso aos dados;	

<sup>8</sup> UK Data Service. 2019. Disponível em: <https://www.youtube.com/watch?v=-wjFMMQD3UA>. Acesso: 18 maio 2022.

	Fomentar dados.
Gestão a longo prazo	Migrar dados para o melhor formato;
	Migrar dados para um meio adequado;
	Fazer <i>backup</i> e armazenar dados;
	Coletar e produzir metadados e documentação;
	Preservar e curar dados.
Reutilização	Realizar análises secundárias;
	Realizar pesquisas de acompanhamento;
	Realizar revisões de pesquisa;
	Examinar as descobertas;
	Usar dados para ensinar e aprender.

Fonte: Elaborado por Almeida (2019, p. 53), com base em Corti *et al.* (2014, p. 17-18).

### 2.3.2 Descrição dos metadados

Os metadados são, basicamente, dados que descrevem outros dados ou, ainda, dados sobre dados, essa é uma clássica e simplificada definição presente na Biblioteconomia. Os metadados são descrições representativas que os bibliotecários estão acostumados a realizarem no trabalho do dia a dia, são descrições dos seus itens, geralmente, bibliográficos, para armazená-los em um banco de dados. Segundo Berman (2008, p. 56),

[...] os metadados são necessários para poder encontrar e usar seus dados imediatamente e também nos próximos anos. Identificar padrões relevantes para conteúdo e formato de dados/metadados, seguindo-os para garantir que os dados possam ser usados por outros [...].

Dessa forma:

Os metadados são dados cujo propósito é garantir a autenticidade, descrever, tornar acessíveis, ou de alguma forma qualificar e aumentar a inteligibilidade dos dados de base. Quando o significado dos dados está acessível ao utilizador menos especializado, como é o caso dos registos de publicações, existem normas bem estabelecidas para os metadados, neste exemplo as diversas normas bibliográficas. Na maioria das áreas que produzem dados científicos, não estão fixadas normas para metadados, o que torna a tarefa de descrição um desafio substancial (RIBEIRO *et al.*, 2010, p. 49).

Tratando-se de itens digitais (não bibliográficos), no caso dos dados de pesquisa, uma boa definição de metadados, segundo Pavão *et al.* (2019, p. 7-8), é de que eles são:

[...] informações sobre um item no repositório, incluindo, mas não limitado, ao nome do autor, data de publicação, título do conteúdo de dados, descrição do conteúdo e outras informações relacionadas.

Alguns exemplos de metadados que precisam ser descritos são: nome do projeto, laboratório de pesquisa, agência de financiamento, equipe, instrumentos utilizados, local, clima,

recursos e produtos, entre outras informações, ou seja, tudo que pode interferir no resultado dos dados.

Os metadados são indispensáveis para a recuperação e reutilização de conjuntos de dados. Eles fornecem informações sobre o criador dos dados, o assunto dos dados, o tipo de arquivo, informações geográficas e outros aspectos. Em outras palavras, os metadados são “dados sobre dados”. Os metadados fazem uso de padrões internacionais para troca de dados. Isso garante que as informações e o conjunto de dados associado possam ser encontrados pelos mecanismos de pesquisa (PAVÃO *et al.*, 2019, p. 9).

A descrição dos metadados é tão importante quanto o seu correto armazenamento e sua preservação a longo prazo. Somente com a descrição correta dos metadados, os dados poderão ser descobertos e reutilizados. A atribuição dos metadados de dados de pesquisa deverá ficar a cargo, principalmente, dos pesquisadores, desde antes da coleta até depois da finalização do armazenamento nos repositórios. Como visto nos ciclos de vida dos dados, tem a etapa da descrição dos dados coletados, ou seja, dos metadados.

Existe uma infinidade de tipos de metadados, alguns gerais e aplicados a quase todas as áreas do conhecimento, como o padrão *Dublin Core* ou *DataCite Metadata Schema* (ver Anexo A). Adotar um padrão de metadados internacional é importante para a interoperabilidade dos dados, para que eles consigam ser migrados e processados por outros sistemas. E para que outros pesquisadores consigam entender os dados e variáveis presentes neles.

Intimamente ligada ao desenvolvimento de padrões na citação de dados, existe uma cada vez maior conscientização de que é necessário conservar, descrever e proporcionar o acesso aos conjuntos de dados de maneira correta. As atividades relacionadas a isso se agrupam no que se denomina *data curation* ou curadoria de dados. Para que um conjunto de dados seja citado, antes deve ter sido arquivado em um repositório, preservado em um formato interoperativo, descrito adequadamente por um grupo formal de metadados conectado ao conjunto de dados e colocado à disposição de outros pesquisadores para sua reutilização (SILVA, 2019, p. 90).

Existem algumas perguntas que ajudam a guiar a escolha dos metadados a serem descritos. Segundo Sayão e Sales (2015, p. 20-21, grifo do autor), são três questionamentos: primeiro: “**Que metadados são necessários?** Neste momento você tem que definir qual é o **elenco de metadados que são necessários** para que os **dados possam ter significado** e possam ser interpretados ao longo do tempo e do espaço.”; segundo: “[...] **como os metadados serão criados e/ou capturados?** Você deve informar também no seu Plano de Gestão de Dados **como os metadados serão criados ou capturados.**”; e, terceiro: “**Que esquema ou padrão de metadado será usado? Informe que esquema (ou formato) de metadados você vai usar para descrever os seus dados**”. Para os autores, deve-se “[...] justificar no Plano a escolha do

esquema de metadados, considerando a **sua comunidade de pesquisa, o repositório ou centro de dados** que você usará para arquivar os dados e a natureza do **seu projeto**”.

Para que as coleções de dados possam ser reusadas de forma confiável é preciso que essas coleções sejam limpas, organizadas, documentadas e descritas por metadados gerais e disciplinares que evidenciem os instrumentos e métodos usados para obtê-las e, finalmente, disponibilizadas em bases de dados para que alguém não envolvido na pesquisa possa acessá-las, interpretá-las e ressignificá-las, posto que os dados não falam por si próprios, como um artigo de periódico. No entanto, em muitos casos as coleções precisam passar por processos custosos de anonimização, ou seja, há um trabalho extra considerado “não científico” e não reconhecido pelos esquemas de recompensa que precisa ser feito para revelar o valor dos dados (SAYÃO; SALES, 2018, p. 4.190).

A definição de qual padrão de metadados será utilizado é importante para a descrição da pesquisa e suas especificidades, para o reuso e para sua preservação no futuro. O pesquisador deve escolher o tipo de padrão de metadados de acordo com cada pesquisa ou tipo de dados, ou seja, não necessariamente adotará para todas as coletas o mesmo padrão. Este poderá ser auxiliado pelo “bibliotecário de dados”, que deverá aprofundar os conhecimentos e particularidades em relação aos tipos e às diferenças em padrões de metadados (Anexo A). Quem mais trabalha com os padrões de metadados são os pesquisadores na fase de coleta dos dados, mas na fase de depósito do conjunto de dados em repositório de dados, o bibliotecário realiza a descrição do conjunto de dados (se o depósito ficar a cargo das bibliotecas), mas em caso de autodepósito pelos pesquisadores, o bibliotecário deve realizar a conferência e a correção dos metadados administrativos (caso necessário) antes de autorizar a publicação do conjunto de dados.

### **2.3.3 Elementos para a citação do conjunto de dados**

A citação dos dados é fundamental aos pesquisadores que vão reusar os conjuntos de dados de outros colegas. Alguns modelos de citação de dados e os elementos necessários para isso estão disponíveis, entretanto, a norma brasileira de referências (NBR 6023), da Associação Brasileira de Normas Técnicas (ABNT), não inclui exemplos, elementos e modelos de como citar conjunto de dados. De acordo com Dudziak (2016, não paginado), “[...] os princípios da citação de dados estabelecem que os dados de pesquisa devem estar facilmente acessíveis, ser devidamente preservados, citados e referenciados”. Os elementos ou campos das citações são também tipos de informações (ou metadados) para se encontrar/localizar o que foi utilizado na composição de um trabalho ou de uma pesquisa científica. A ferramenta DMPTool (2022) cita cinco elementos principais para a citação de conjunto de dados e cinco elementos adicionais comuns:

Quadro 7 – Elementos principais e elementos adicionais de citação DMPTool

	<b>Elementos principais</b>	<b>Elementos adicionais</b>
1	Criadores que podem ser indivíduos ou organizações.	Versão do conjunto de dados analisado no artigo de citação.
2	Título do projeto de pesquisa.	Data de acesso quando os dados foram acessados para análise no artigo citante.
3	Ano de publicação quando o conjunto de dados foi lançado (pode ser diferente da data de acesso).	Subconjunto do conjunto de dados analisados (por exemplo, um intervalo de datas ou números de registro, uma lista de variáveis).
4	Editor (o <i>data center</i> , arquivo ou repositório).	Verificador de que o conjunto de dados ou subconjunto acessado por um leitor é idêntico ao analisado pelo autor (por exemplo, um Checksum).
5	Identificador – um identificador público único (por exemplo, um ARK ou DOI).	Localização do conjunto de dados na internet, necessária se o identificador não for “acionável” (conversível em um endereço da <i>web</i> ).

Fonte: Adaptado de DMPTool (2022).

### 2.3.4 O Plano de Gestão de Dados (PGD)

O Plano de Gestão de Dados (PGD) é o documento formal que descreve o ciclo de vida dos dados. É um documento que registra as decisões tomadas, as ações que ocorreram, o que foi usado de equipamentos, de materiais e de métodos, ou seja, registra tudo o que aconteceu para se chegar àqueles dados. Segundo Silva (2019, p. 55), “[...] registra as decisões tomadas em relação aos metadados, formatos, bases de dados, métodos, segurança e períodos de armazenamento, assim como os custos associados com a gestão de dados”.

O Plano de Gestão de Dados (PGD) é um documento formal relacionado ao projeto de pesquisa que deve responder a duas perguntas básicas: **Quais dados são gerados?** O pesquisador deve informar o formato dos dados e uma breve descrição dos mesmos, de tal forma que aqueles que pretendem utilizar possam compreendê-los. Os dados gerados podem, por exemplo, estar em formato de planilha eletrônica, texto ou bancos de dados digitais. **Como e onde esses dados são armazenados e disponibilizados?** O pesquisador deve informar o local de armazenamento e disponibilização dos dados, que pode ser um repositório público ou particular do pesquisador. A resposta a esta pergunta deve envolver aspectos éticos e legais que eventualmente estão envolvidos nos dados (UNIVERSIDADE DE SÃO PAULO, 2020, não paginado, grifo nosso).

Portanto, o PGD é a consolidação das decisões tomadas de como proceder com o tratamento dos dados, desde a fase do planejamento da pesquisa até a preservação de dados, passando pelos direitos de propriedade intelectual etc. Conforme DMPTool (2022), o PGD descreve o que você fará com seus dados durante e após uma pesquisa, assim, você documenta e evita problemas, planejando o processo e a estrutura correta, o que não leva muito tempo e traz benefícios a longo prazo.

O PGD descreve o ciclo de vida de gestão para todos os dados que serão coletados, processados ou gerados por um projeto de pesquisa. De uma forma abreviada, ele se

constitui em um documento formal que estabelece um compromisso de como esses dados serão tratados durante todo o desenvolvimento do projeto, e também após a sua conclusão. Para isso, o PGD descreve, de uma forma geral, que dados serão processados, coletados ou gerados; quais as metodologias e padrões que serão utilizados nesses processos; se, como e sob que condições esses dados serão compartilhados e/ou tornados abertos para a comunidade de pesquisa; e como eles serão curados e preservados (SAYÃO; SALES, 2015, p. 15).

Segundo Sayão e Sales (2015), o PGD não é um documento burocrático, ele precisa deixar de ser uma simples tarefa administrativa, realizada somente porque é exigência. Deve se perder o caráter generalista, sempre buscando um documento mais específico para aqueles dados, rico em detalhes e descrições e não um texto padronizado a ser usado para todos os projetos. Ainda de acordo com os autores (SAYÃO; SALES, 2015, p. 15, grifo do autor), mesmo sendo um documento **“Ele deve ser tratado como uma carta de intenções que considere o que realmente é necessário para a preservação, compartilhamento e reuso dos dados”**. Para que isso aconteça, é necessário informações sobre os dados, é preciso conseguir responder às perguntas pré-determinadas (Quadro 8), quase que como um roteiro para o sucesso da curadoria dos dados.

Quadro 8 – Informações sobre dados de pesquisa

<b>INFORMAÇÕES SOBRE OS DADOS</b>	
<b>Que tipo de dados sua pesquisa vai produzir?</b>	<b>Liste os dados que seu projeto irá produzir e os caracterize</b> em termos de natureza, origem e processamento: eles podem ser observacionais, experimentais, brutos ou derivados, simulações, coleções físicas, modelos, <i>software</i> , imagens, vídeos e muito mais.
<b>Que quantidade de dados será gerada pela pesquisa?</b>	Com base na sua hipótese e no plano de amostragem, avalie o <b>volume de dados</b> que o seu projeto irá gerar.
<b>Como os dados serão coletados?</b>	Você deve especificar também os <b>métodos como os dados serão adquiridos</b> , isso inclui informações sobre quem, o que, quando e onde (como as amostras serão coletadas e analisadas? Que instrumentos serão usados?).
<b>Como os dados serão processados?</b>	Uma vez que os dados foram adquiridos, deve ser especificado como eles serão processados. “Esta etapa deve ser considerada antes do projeto, pois pode afetar a maneira como os dados serão organizados, quais formatos serão usados, e quanto deve ser previsto, em termos orçamentários, para <i>hardware</i> e <i>software</i> . Devem ser considerados neste momento itens como que <i>software</i> poderá ser usado, que algoritmos serão empregados, e como esses itens se enquadram no fluxo de trabalho do projeto” (STRASSER, 2012).
<b>Quais os formatos de arquivo que serão usados?</b>	Os formatos de arquivo dos dados que você planeja usar devem ser declarados e sua escolha deve ser justificada. Descreva os formatos nas fases de submissão, distribuição e preservação, observando que esses formatos podem ser os mesmos. Na sua escolha, você deve considerar os padrões que são usados na sua área de pesquisa. Se os dados forem arquivados por longo prazo, é necessário considerar o uso de formatos padronizados e não proprietários, que são mais fáceis de serem interpretados no futuro, de forma independente de plataforma tecnológica ( <i>hardware</i> e <i>software</i> ). (Considere se um banco de dados relacional ou outra estratégia de organização de dados pode ser mais apropriado para a sua pesquisa).

Como os arquivos serão nomeados?	É importante descrever também a convenção <b>adotada para dar nomes para os seus conjuntos de dados, arquivos e pastas</b> . Convencionando isso de antemão, você estará menos propenso a mudar ou reorganizar os arquivos durante o projeto.
Quais são as medidas de garantia e de controle de qualidade?	Você deve identificar quais são as medidas que você planeja adotar para <b>garantir e controlar a qualidade dos dados</b> ; é necessário incluir também o que será feito durante e depois dos dados coletados e, ainda, no curso da análise dos dados.
Há coleções de dados disponíveis que servem para sua pesquisa?	A <b>revisão dos dados existentes em periódicos e arquivos de dados</b> da sua área de pesquisa reforçará o valor de seu projeto e justificará mais claramente porque os dados atualmente disponíveis são inadequados para responder as suas questões de pesquisa.
Serão usados dados já existentes?	Se <b>dados já existentes podem ser usados na sua pesquisa</b> , identificá-los e determinar suas origens (proveniência) é uma informação importante e deve ser registrada, bem como, a relação entre esses dados e os dados que você está coletando. Se a sua coleção de dados será combinada com dados já existentes, cabe definir como será <b>assegurada a compatibilidade de formatos</b> .
Como os dados serão mantidos a curto prazo?	Você precisa descrever <b>como os dados serão gerenciados logo após o término do projeto</b> ; isso significa planejar como manter o controle sobre as diferentes versões dos seus dados e das análises; como você fará <i>backup</i> de seus dados; se há computadores destinados a isso na sua instituição. Considere as opções de <i>backup</i> na sua instituição ( <i>on-site</i> ) e externamente ( <i>off-site</i> ). Descreva a sua estratégia para garantir a segurança dos dados, especialmente no caso de <b>dados sensíveis</b> . Delineie os possíveis usuários dos dados.
Quem será o responsável pela gestão de curto prazo?	Identifique quem são os responsáveis pela gestão de curto prazo na sua instituição; determine papéis e responsabilidades para a gestão, o arquivamento, o controle de versões e os procedimentos de <i>backup</i> .

Fonte: Adaptado de Sayão e Sales (2015, p. 18-19).

A importância de um Plano de Gestão de Dados (PGD) é ter um documento que coloque por escrito todos os passos para se trabalhar com os dados e para que os inúmeros cientistas da organização ou universidade consigam adaptá-los a sua realidade, independente da área. Os dados precisam ser gerenciados a curto, médio e longo prazo para sua manutenção, integridade e preservação. No curto prazo, os dados ainda estão sob a responsabilidade de quem os produziu, são dados que estão sendo utilizados em pesquisas em curso, portanto, quem faz a gestão são os próprios pesquisadores dentro de seus próprios laboratórios e salas de trabalho localizados de instituições de ensino e de pesquisa. Os dados a médio e longo prazo, geralmente, já foram depositados em repositórios de dados, assim, sua gestão é compartilhada.

Há muitas razões para que seja necessária a elaboração de um plano de gestão de dados, porém a mais importante é que a gestão de dados é uma das áreas essenciais na **conduta responsável da pesquisa nos ambientes científicos atuais**. Além do mais, auxilia os pesquisadores a considerar, **ainda na fase de concepção e planejamento do projeto de pesquisa**, como os dados serão **geridos durante a pesquisa** e como serão **posteriormente preservados e compartilhados** com a comunidade científica mais ampla (EYNDEN *et al.*, 2011 apud SAYÃO; SALES, 2015, p. 16, grifo do autor).

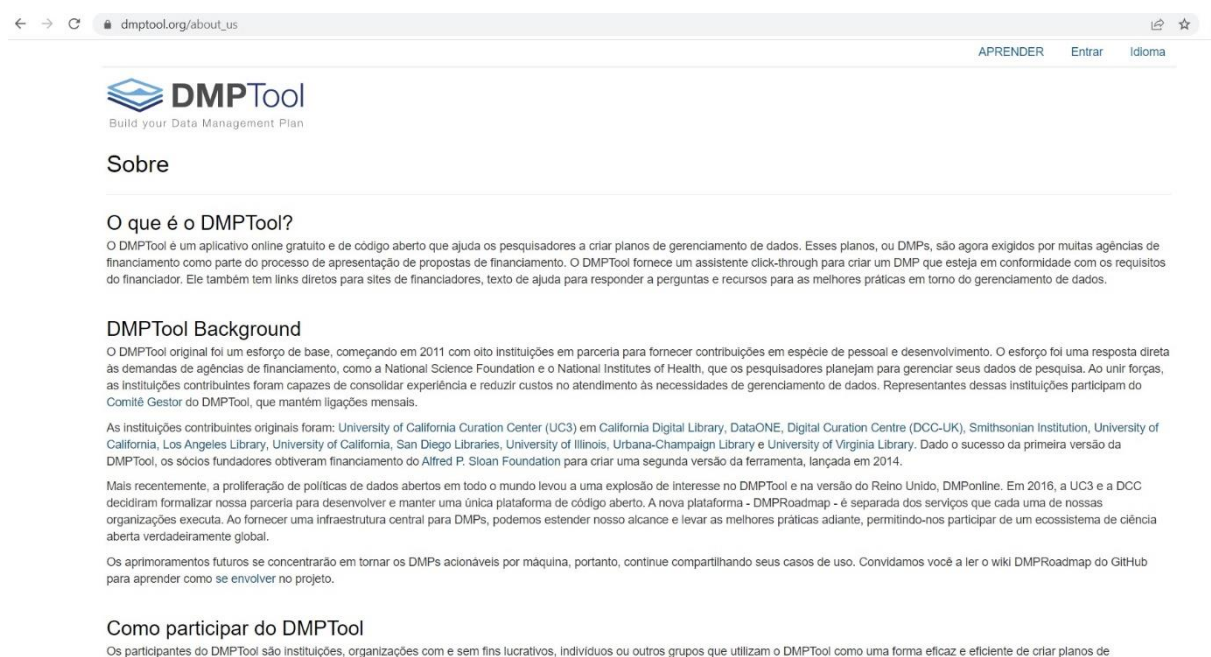


Existem muitos fatores que devem constar num PGD, como custos de armazenamento, orçamento em geral, questões de ética e dados sensíveis e outras pautas. Duas ferramentas<sup>9</sup> para a construção do Plano de Gestão de dados de pesquisa são: DMPTool (<https://dmptool.org/>) e DMPonline (<https://dmponline.dcc.ac.uk/>).

#### 2.3.4.1 A ferramenta DMPTool

O DMPTool é uma ferramenta online utilizada para construir PGDs. Segundo o DMPTool, devem ser descritos os metadados: de identificação (título, criador, identificador, data, método, processamento, fonte e agência financiadora); de descrição de conteúdo (assunto, local, idioma, lista de variáveis e lista de código); de descrição técnica (inventário de arquivos, formato de arquivo, estrutura do arquivo, versão, soma de verificação, *software* necessário, direitos); e de acesso (direitos e informações de acesso). Para mais informações, ver Quadro 9. A ferramenta possui a tradução para o inglês (EUA) e para o português (Brasil) na própria página online.

Figura 6 – Página DMPTool



Fonte: DMPTool (2022).

<sup>9</sup> maDMP – constitui uma tecnologia mais moderna para a construção de planos de gestão de dados, não explorado neste trabalho. Segundo o DataCite (2022, não paginado) Os DMPs acionáveis por máquina (maDMPs) levam o conceito de DMP adiante usando PIDs e serviços PIDs para conectar todos os recursos associados a um DMP.

Em sua documentação, a ferramenta define o que são os termos e explica mais sobre eles, isso para que os pesquisadores se orientem por essas definições e utilizem a ferramenta de maneira adequada.

Quadro 9 – Definições dos metadados do DMPTool

<b>Metadados de identificação</b>	
<b>Título</b>	Nome do conjunto de dados ou projeto de pesquisa que o produziu.
<b>Criador</b>	Nomes e endereços das organizações ou pessoas que criaram os dados; o formato preferido para nomes pessoais é o sobrenome primeiro (por exemplo, Smith, Jane).
<b>Identificador</b>	Número único usado para identificar os dados, mesmo que seja apenas um número de referência interno do projeto.
<b>Data</b>	Datas importantes associadas aos dados, incluindo: data de início e término do projeto; data de lançamento; período coberto pelos dados; e outras datas associadas à vida útil dos dados, como ciclo de manutenção, programação de atualização; o formato preferido é aaaa-mm-dd ou aaaa.mm.dd-aaaa.mm.dd para um intervalo.
<b>Método</b>	Como os dados foram gerados, listando equipamentos e <i>softwares</i> usados (incluindo números de modelo e versão), fórmulas, algoritmos, protocolos experimentais e outras informações que podem ser incluídas em um caderno de laboratório.
<b>Processamento</b>	Como os dados foram alterados ou processados (por exemplo, normalizados).
<b>Fonte</b>	Citações a dados derivados de outras fontes, incluindo detalhes de onde os dados de origem são mantidos e como foram acessados.
<b>Financiador</b>	Organizações ou agências que financiaram a pesquisa.
<b>Descrição do conteúdo</b>	
<b>Assunto</b>	Palavras-chave ou frases que descrevem o assunto ou conteúdo dos dados.
<b>Local</b>	Todos os locais físicos aplicáveis.
<b>Idioma</b>	Todos os idiomas usados no conjunto de dados.
<b>Lista de variáveis</b>	Todas as variáveis nos arquivos de dados, quando aplicável.
<b>Lista de códigos</b>	Explicação dos códigos ou abreviações usados nos nomes dos arquivos ou nas variáveis nos arquivos de dados (por exemplo, “999 indica um valor ausente nos dados”).
<b>Descrição técnica</b>	
<b>Inventário de arquivos</b>	Todos os arquivos associados ao projeto, incluindo extensões (por exemplo, “NWPalaceTR.WRL”, “stone.mov”).
<b>Formatos de arquivo</b>	Formatos dos dados, por exemplo, FITS, SPSS, HTML, JPEG etc.
<b>Estrutura do arquivo</b>	Organização do(s) arquivo(s) de dados e <i>layout</i> das variáveis, quando aplicável.
<b>Versão</b>	Carimbo de data/hora exclusivo e identificador para cada versão.
<b>Soma de verificação</b>	Um valor de resumo calculado para cada arquivo que pode ser usado para detectar alterações; se um resumo recalculado for diferente do resumo armazenado, o arquivo deve ter mudado.
<b>Software necessário</b>	Nomes de quaisquer pacotes de <i>software</i> para fins especiais necessários para criar, visualizar, analisar ou usar os dados de outra forma.
<b>Metadados de acesso</b>	
<b>Direitos</b>	Quaisquer direitos de propriedade intelectual, direitos legais, licenças ou restrições conhecidas sobre o uso dos dados.
<b>Informações de acesso</b>	Onde e como seus dados podem ser acessados por outros pesquisadores.

Fonte: Adaptado de DMPTool<sup>10</sup> (2022).

<sup>10</sup> DMPTOOL.ORG (2022). Disponível em: [https://dmptool.org/general\\_guidance](https://dmptool.org/general_guidance). Acesso: 29 mar. 2022.

Figura 7 – Ferramenta DMPTool

The screenshot displays the DMPTool web interface for creating a new research result. The page title is "Descrevendo os metadados de dados de pesquisa nos Planos de Gestão de Dados". The navigation menu includes "Detalhes do Projeto", "Colaboradores", "Escrever Plano", "Resultados da pesquisa", "Baixar", and "Finalizar / Publicar". The user is identified as "Igor Matos" and the language is set to "Idioma".

The main form, titled "Novo resultado de pesquisa", contains the following elements:

- Tipo:** A dropdown menu with the placeholder text "- Por favor escolha um -".
- Título:** A text input field.
- Abreviação:** A text input field.
- Descrição:** A rich text editor with a toolbar containing icons for bold, italic, list, link, and image.
- Confidentiality and Personal Information:** Two checkboxes: "Pode conter dados confidenciais?" and "Pode conter informações de identificação pessoal?".
- Repositórios pretendidos:** A section with a button "Adicionar um repositório".
- Padrões de metadados:** A section with a button "Adicionar um padrão de metadados".
- Data de lançamento prevista:** A date input field with the format "dd / mm / aaaa".
- Nível de acesso inicial:** A dropdown menu with "Open" selected.
- Licença inicial:** A dropdown menu with the placeholder text "- Por favor escolha um -".
- License Information:** A link to "https://creativecommons.org/about/licenses/" for more information.
- Buttons:** "Salvar" and "Cancelar" buttons at the bottom of the form.

The footer includes the CDLI logo, contact information, and a copyright notice: "© 2022 Os Regents da Universidade da Califórnia".

Fonte: DMPTool (2022).

O DMPTool permite acrescentar resultados de pesquisas, caso tenha mais de uma publicação para aquele conjunto de dados. A opção “tipo” possibilita escolher: audiovisual, coleção, documento de dados, conjunto de dados, evento, imagem, recurso interativo, representação do modelo, objeto físico, serviços, programa, som, texto, fluxo de trabalho e outros. Há, ainda, os outros campos como o título e a descrição (resumo explicativo daquele resultado), o repositório pretendido, que abre uma nova caixa em que se deve preencher qual área do assunto (ex. Química, Biologia, Medicina, Engenharia de Construção, Arquitetura etc.), o tipo de repositório (multidisciplinar, disciplina específica ou institucional) e os termos do assunto (palavras-chave).

#### 2.3.4.2 A ferramenta DMPonline

O DMPonline é uma ferramenta que permite diferentes descrições e customizações, possibilitando, inclusive, criar um modelo personalizado para as instituições e disponibilizá-las como *template* na própria página da ferramenta. É possível também compartilhar com outros pesquisadores a construção do PGD e escolher qual a visibilidade do plano: privado (restrito a

você e seus colaboradores), organizacional (qualquer pessoa em sua organização pode visualizar seu plano), ou público (qualquer pessoa pode visualizar seu plano), os dois últimos serão visualizados pelas pessoas autorizadas, mas somente após a publicação e a conclusão do plano. Para publicar qualquer PGD, é necessário preencher pelo menos 50% de todos os campos. O DMPonline encontra-se disponível nos idiomas: alemão, inglês (UK), inglês (EUA), espanhol e francês.

Figura 8 – Ferramenta DMPonline

### Descrevendo os metadados de dados de pesquisa nos Planos de Gestão de Dados

The screenshot shows the 'Detalhes do Projeto' (Project Details) section of the DMPonline interface. The form is divided into several sections:

- \*Título do projeto:** A text input field containing 'Descrevendo os metadados de dados de pesquisa nos Planos de'.
- Checkboxes:** A checked checkbox labeled 'projeto simulado para teste, prática ou fins educacionais'.
- Resumo do projeto:** A rich text editor with a toolbar containing bold (B), italic (I), bulleted list, numbered list, link, and table icons.
- Início do projeto:** A date input field with the placeholder 'dd/mm/aaaa' and a calendar icon.
- Fim do projeto:** A date input field with the placeholder 'dd/mm/aaaa' and a calendar icon.
- identificação:** A text input field containing '97187'.
- Financiador:** A text input field with the placeholder 'Comece a digitar para ver uma lista de sugestões'.
- Status do financiamento:** A dropdown menu with the selected option '- Por favor escolha um -'.
- Número/url da concessão:** A text input field.

On the right side, there is a section titled 'Selecionar orientação' (Select orientation). It includes a sub-header 'Para ajudá-lo a escrever seu plano, o DMPonline pode mostrar orientações de várias organizações.' and a sub-section 'Selecione até 6 organizações para ver suas orientações.' with a checked checkbox for 'Centro de Curadoria Digital'. Below this, it says 'Encontre orientação de outras organizações abaixo' and 'Veja a lista completa'. A 'Salvar' (Save) button is located at the bottom right of the form.

Fonte: DMPonline (2022).

As estruturas das duas ferramentas são similares (inclusive em algumas exigências), principalmente nas perguntas e no formato do plano, na visibilidade etc. As maiores diferenças estão no número de línguas que podem ser traduzidas e nos resultados de pesquisa que a ferramenta DMPTool permite vincular a mais em seus PGDs. Nos demais aspectos, são plataformas que conseguem descrever o PGD, o que permite que sejam utilizadas e disseminadas para uso no Brasil, sem a necessidade de recriar uma ferramenta nacional para isso.

Quadro 10 – Estrutura de metadados das ferramentas DMPToll e DMPonline

<b>Coleção de dados</b>	
<b>Perguntas gerais</b>	<b>Questões a considerar</b>
Quais dados você coletará ou criará?	<ol style="list-style-type: none"> <li>1. Qual o tipo, o formato e o volume de dados?</li> <li>2. Os formatos e <i>softwares</i> escolhidos permitem o compartilhamento e o acesso de longo prazo aos dados?</li> <li>3. Existem dados existentes que você pode reutilizar?</li> </ol>
Como os dados serão coletados ou criados?	<ol style="list-style-type: none"> <li>1. Quais padrões ou metodologias você usará?</li> <li>2. Como você estruturará e nomeará suas pastas e arquivos?</li> <li>3. Como você vai lidar com o versionamento?</li> <li>4. Quais processos de garantia de qualidade você adotará?</li> </ol>
<b>Documentação e metadados</b>	
Que documentação e metadados acompanharão os dados?	<ol style="list-style-type: none"> <li>1. Quais informações são necessárias para que os dados sejam lidos e interpretados no futuro?</li> <li>2. Como você irá capturar/criar esta documentação e estes metadados?</li> <li>3. Quais padrões de metadados você usará e por quê?</li> </ol>
<b>Ética e conformidade legal</b>	
Como você vai gerenciar quaisquer questões éticas?	<ol style="list-style-type: none"> <li>1. Você obteve consentimento para preservação e compartilhamento de dados?</li> <li>2. Como você protegerá a identidade dos participantes, se necessário? Por exemplo, através de anonimização.</li> <li>3. Como os dados confidenciais serão tratados para garantir que sejam armazenados e transferidos com segurança?</li> </ol>
Como você gerenciará questões de direitos autorais e direitos de propriedade intelectual (DPI)?	<ol style="list-style-type: none"> <li>1. Quem é o proprietário dos dados?</li> <li>2. Como os dados serão licenciados para reutilização?</li> <li>3. Existem restrições à reutilização de dados de terceiros?</li> <li>4. O compartilhamento de dados será adiado/restringido, por exemplo, para publicar ou buscar patentes?</li> </ol>
<b>Armazenamento e backup</b>	
Como os dados serão armazenados e copiados durante a pesquisa?	<ol style="list-style-type: none"> <li>1. Você tem armazenamento suficiente ou precisará incluir cobranças por serviços adicionais?</li> <li>2. Como será feito o <i>backup</i> dos dados?</li> <li>3. Quem será o responsável pelo <i>backup</i> e pela recuperação?</li> <li>4. Como os dados serão recuperados no caso de um incidente?</li> </ol>
Como você gerenciará o acesso e a segurança?	<ol style="list-style-type: none"> <li>1. Quais são os riscos para a segurança dos dados e como eles serão gerenciados?</li> <li>2. Como você controlará o acesso para manter os dados seguros?</li> <li>3. Como você garantirá que os colaboradores possam acessar seus dados com segurança?</li> <li>4. Se estiver criando ou coletando dados em campo, como você garantirá a transferência segura para seus principais sistemas protegidos?</li> </ol>
<b>Seleção e preservação</b>	
Quais dados são de valor a longo prazo e devem ser retidos, compartilhados e/ou preservados?	<ol style="list-style-type: none"> <li>1. Quais dados devem ser retidos/destruídos para fins contratuais, legais ou regulatórios?</li> <li>2. Como você decidirá quais outros dados manter?</li> <li>3. Quais são os usos previsíveis de pesquisa para os dados?</li> <li>4. Por quanto tempo os dados serão retidos e preservados?</li> </ol>
Qual é o plano de preservação a longo prazo para o conjunto de dados?	<ol style="list-style-type: none"> <li>1. Onde, por exemplo, em qual repositório ou arquivo os dados serão mantidos?</li> <li>2. Quais custos, se houver, serão cobrados pelo repositório ou arquivo de dados selecionado?</li> <li>3. Você gastou tempo e esforço para preparar os dados para compartilhamento/preservação?</li> </ol>
<b>Compartilhamento de dados</b>	
Como você compartilhará os dados?	<ol style="list-style-type: none"> <li>1. Como os usuários em potencial descobrirão seus dados?</li> <li>2. Com quem você compartilhará os dados e em que condições?</li> <li>3. Você compartilhará dados por meio de um repositório, lidará com solicitações diretamente ou usará outro mecanismo?</li> </ol>

	<ol style="list-style-type: none"> <li>4. Por quando você disponibilizará os dados?</li> <li>5. Você pretende obter um identificador persistente para seus dados?</li> </ol>
Alguma restrição no compartilhamento de dados é necessária?	<ol style="list-style-type: none"> <li>1. Que ação você tomará para superar ou minimizar as restrições?</li> <li>2. Por quanto tempo você precisa de uso exclusivo dos dados e por quê?</li> <li>3. Será necessário um contrato de compartilhamento de dados (ou equivalente)?</li> </ol>
<b>Responsabilidades e recursos</b>	
Quem será responsável pelo gerenciamento de dados?	<ol style="list-style-type: none"> <li>1. Quem é o responsável por implementar o DMP e garantir que ele seja revisado?</li> <li>2. Quem será responsável por cada atividade de gerenciamento de dados?</li> <li>3. Como as responsabilidades serão divididas entre os sites parceiros em projetos de pesquisa colaborativa?</li> <li>4. A propriedade dos dados e as responsabilidades pelo RDM farão parte de algum acordo de consórcio ou contrato acordado entre os parceiros?</li> </ol>
Quais recursos você precisará para entregar seu plano?	<ol style="list-style-type: none"> <li>1. É necessária especialização adicional (ou treinamento para o pessoal existente)?</li> <li>2. Você precisa de <i>hardware</i> ou <i>software</i> que seja adicional ou excepcional à provisão institucional existente?</li> <li>3. As cobranças serão aplicadas pelos repositórios de dados?</li> </ol>

Fonte: Adaptado de DMPTool e DMPonline (2022).

### 2.3.5 Tipos de licenças

O tipo de licença comercial geralmente atribuída a livros é o *Copyright*, porém, para os dados de pesquisa depositados em repositórios, recomendam-se as licenças abertas. Segundo Silva (2019, p. 37), “[...] ainda quando os dados estão sob os direitos do autor, é recomendável uma licença aberta, como a *Creative Common (CC)* ou *Open Data Commons* para compartilhar”.

Denominadas de licenças *Copyleft*, são licenças abertas que autorizam o uso das obras com base nos direitos de propriedade intelectual e permitem a livre circulação, o acesso e a modificação, sendo necessário que obras derivadas utilizem do mesmo tipo de licença para que os seus conteúdos permaneçam abertos e democráticos.

A licença *Copyright* é o tipo mais comum utilizado pela indústria editorial para obras comerciais, pois restringem o uso e a comercialização da obra, que deve ser adquirida/comprada para que o usuário obtenha esse direito de uso. No *Copyright*, os autores passam os direitos de comercialização de suas obras para as editoras.

O domínio público é o uso irrestrito da obra, ou seja, não possui nenhuma licença ou proteção comercial. Pela legislação brasileira, uma obra cai em domínio público 70 anos após a morte do autor, o prazo começa a contar em janeiro do ano posterior à morte. Nas licenças *Creative Commons* e *Open Data Commons* existem tipos de licenças em domínio público, não sendo necessário seguir a legislação (70 anos após a morte do autor) criada para resguardar a licença comercial e a exploração por parte da indústria editorial.




Conforme mencionado por Seabra (2014, p. 78), “[...] os direitos autorais são inalienáveis e irrenunciáveis, e, entre eles, inscrevem-se o de autoria, que é eterno; o de integridade da obra, e o conseqüente direito de modificá-la [...]”. Mesmo que uma obra esteja em domínio público, é fundamental a sua citação e o apontamento dos devidos créditos aos autores, quem infringir essa regra pode responder por plágio e ter o título cassado, se for em graduação, mestrado ou doutorado.




### 2.3.5.1 As licenças *Creative Commons*

As *Creative Commons* são licenças criadas principalmente por conta do Movimento de Acesso Aberto (MAA), elas possuem diferentes tipos de permissões, que concedem o direito ao uso aberto, somente necessitando dar os devidos créditos aos seus autores e não infringir à finalidade que a licença permite. Segundo Leite *et al.* (2012), as licenças *Creative Commons* são aquelas que o próprio autor, sem necessitar de intermediários (como advogados), atribui à sua obra. Ao utilizar essas licenças, o autor informa ao mundo o que ele permite que terceiros o utilizem, sem que estes tenham que lhe pedir autorização.

As licenças *Creative Commons* são utilizadas pela Open Archives Initiative (OAI). No modelo OAI, as licenças *Creative Commons* destinam-se a declarar os usos permitidos dos objetos digitais e foram estudadas para comunicar aos usuários finais, com os patrocinadores e com as máquinas. Consistem num esquema de metadados e são utilizados junto com uma linguagem de máquina denominada ODRL (*Open Digital Rights Language*) (TAMMARO, 2008, p. 289).

Quadro 11 – Licenças *Creative Commons*

Tipo de licença	Descrição da licença
 <p data-bbox="359 1563 478 1619">Atribuição CC BY</p>	<p data-bbox="614 1464 1452 1619">Essa licença permite que outras pessoas distribuam, remixem, adaptem e desenvolvam seu trabalho, mesmo comercialmente, desde que sejam creditadas pela criação original. Essa é a licença mais flexível oferecida. Recomendado para máxima disseminação e uso de materiais licenciados.</p>
 <p data-bbox="263 1765 574 1821">Atribuição-CompartilhaIgual CC BY-SA</p>	<p data-bbox="614 1655 1452 1832">Esta licença permite que outras pessoas remixem, adaptem e desenvolvam seu trabalho, mesmo para fins comerciais, desde que creditem você e licenciem suas novas criações sob os mesmos termos. Essa licença geralmente é comparada às licenças de <i>software</i> livre e de código aberto “<i>copyleft</i>”. Todos os novos trabalhos baseados no seu terão a mesma licença, portanto, quaisquer derivados também permitirão uso comercial.</p>
 <p data-bbox="271 1951 566 2007">Atribuição-SemDerivações CC BY-ND</p>	<p data-bbox="614 1872 1452 1989">Esta licença permite que outras pessoas reutilizem o trabalho para qualquer finalidade, inclusive, comercialmente; no entanto, não pode ser compartilhado com outras pessoas de forma adaptada e o crédito deve ser fornecido a você.</p>

 <p>Atribuição-NãoComercial CC BY-NC</p>	<p>Essa licença permite que outras pessoas remixem, adaptem e desenvolvam seu trabalho não comercialmente, e embora os novos trabalhos deles também reconheçam você e não sejam comerciais, eles não precisam licenciar seus trabalhos derivados nos mesmos termos.</p>
 <p>Atribuição-NãoComercial- CompartilhaIgual CC BY-NC-SA</p>	<p>Esta licença permite que outras pessoas remixem, adaptem e desenvolvam seu trabalho de forma não comercial, desde que creditem a você e licenciem suas novas criações sob os mesmos termos.</p>
 <p>Atribuição-NãoComercial- SemDerivações CC BY-NC-ND</p>	<p>Esta é a mais restritiva entre as seis licenças principais, só permitindo que outros façam <i>download</i> dos seus trabalhos e os compartilhem desde que lhe sejam atribuídos os devidos créditos, mas sem que possam alterá-los de nenhuma forma ou utilizá-los para fins comerciais.</p>

Fonte: Adaptado de *Creative Commons* (2021).

### 2.3.5.2 *Open Data Commons*

O *Open Data Commons* é mantido pela Open Knowledge Foundation, uma organização sem fins lucrativos, constituída na Inglaterra e no País de Gales. A Open Knowledge Foundation tem por missão “um mundo aberto, onde todas as informações não pessoais são abertas, gratuitas para que todos possam usar, construir e compartilhar; e criadores e inovadores são reconhecidos e recompensados de forma justa”. Os três tipos de licenças *Open Data* são:

- *Open Data Commons Open Database License* (ODbL) – atribuição compartilhada da mesma forma para dados/bancos de dados;
- *Open Data Commons Attribution License* (ODC-By) – atribuição para dados/bancos de dados;
- *Open Data Commons Public Domain Dedication and License* (PDDL) – domínio público para dados/bancos de dados.

A *Open Data Commons Attribution License* (ODC-By) é um contrato de licença destinado a permitir que os usuários compartilhem, modifiquem e usem livremente para qualquer finalidade e sem quaisquer restrições. Essa licença destina-se ao uso em bancos de dados ou seus conteúdos (“dados”), em conjunto ou individualmente. É específica para a reutilização de banco de dados e conjuntos de dados, assim, pode ser mais apropriada que a *Creative Commons* para os dados de pesquisa.



O *Open Data Commons – Public Domain Dedication and License* (PDDL) é um documento destinado a permitir que você compartilhe, modifique e use livremente este trabalho para qualquer finalidade e sem quaisquer restrições. Esta licença destina-se ao uso em bancos de dados ou seus conteúdos (“dados”), em conjunto ou individualmente.

Quadro 12 – Licenças *Open Data Commons*

NOME DA LICENÇA	TIPO DE LICENÇA	TRECHOS IMPORTANTES DA LICENÇA
<p><i>OPEN DATA COMMONS ATTRIBUTION LICENSE</i> (ODC-BY)</p> <p>(<a href="https://opendatacommons.org/licenses/by/1-0/">https://opendatacommons.org/licenses/by/1-0/</a>)</p>	<p>Atribuição para dados/bancos de dados</p>	<p>A <i>Open Data Commons Attribution License</i> é um contrato de licença destinado a permitir que os usuários compartilhem, modifiquem e usem livremente este banco de dados, sujeito apenas aos requisitos de atribuição estabelecidos na Seção 4.</p> <p>[...]</p> <p>Os bancos de dados podem conter uma grande variedade de tipos de conteúdo (imagens, material audiovisual, sons, todos no mesmo banco de dados, por exemplo), e, portanto, esta licença rege apenas os direitos sobre o banco de dados, e não o conteúdo do banco de dados individualmente. Os licenciadores podem, portanto, querer usar esta licença juntamente com outra licença para o conteúdo.</p> <p>[...]</p> <p>Sujeito aos termos e condições desta licença, o licenciador concede a você uma licença mundial, isenta de royalties, não exclusiva, rescindível (mas apenas sob a Seção 9) para usar o banco de dados pela duração de quaisquer direitos autorais e direitos de banco de dados aplicáveis. Esses direitos incluem explicitamente o uso comercial e não excluem nenhum campo de atuação. Na medida do possível na jurisdição relevante, esses direitos podem ser exercidos em todas as mídias e formatos, sejam agora conhecidos ou criados no futuro.</p>
<p><i>OPEN DATA COMMONS PUBLIC DOMAIN DEDICATION AND LICENSE</i> (PDDL)</p> <p>(<a href="https://opendatacommons.org/licenses/pddl/1-0/">https://opendatacommons.org/licenses/pddl/1-0/</a>)</p>	<p>Domínio público para dados/bancos de dados</p>	<p>Dedicação, renúncia e licença de direitos autorais e direitos de banco de dados</p> <p>Dedicação de direitos autorais e de banco de dados ao domínio público. O titular de direitos, ao usar este documento, dedica a obra ao domínio público para benefício do público e renuncia a todos os direitos de direitos autorais e de banco de dados sobre a obra</p>
<p><i>OPEN DATABASE LICENSE</i> (ODBL)</p> <p>(<a href="https://opendatacommons.org/licenses/odbl/1-0/">https://opendatacommons.org/licenses/odbl/1-0/</a>)</p>	<p>Atribuição compartilhada da mesma forma para dados/bancos de dados</p>	<p>A <i>Open Database License</i> (ODbL) é um contrato de licença destinado a permitir que os usuários compartilhem, modifiquem e usem livremente este banco de dados, mantendo essa mesma liberdade para outros. Muitos bancos de dados são protegidos por direitos autorais e, portanto, este documento licencia esses direitos.</p> <p>[...]</p> <p>Os bancos de dados podem conter uma grande variedade de tipos de conteúdo (imagens, material audiovisual, sons, todos no mesmo banco de dados, por exemplo), e assim o ODbL rege apenas os direitos sobre o banco de dados, e não o conteúdo do banco de dados individualmente.</p>

Fonte: Elaborado a partir das licenças Open Data Commons (2022).

## 2.4 O SERVIÇO REPOSITÓRIO DE DADOS: ANTECEDENTES

Para a contextualização dos antecedentes dos repositórios de dados, é necessário voltar ao contexto das bibliotecas digitais que armazenam arquivos/objetos em dígitos binários, ou seja, zero e um. A discussão do que é uma biblioteca digital já foi em muito superada, debatida em publicações e em eventos, nos anos finais do século XX e na primeira década do século XXI, com inúmeros eventos brasileiros<sup>11</sup> para entender e aprender sobre esse tipo de biblioteca. O modo de organização é um elemento importante para uma biblioteca digital, conforme Tammaro (2008, p. 119), uma “[...] definição, surgida no mundo da ciência da informação, identifica três componentes essenciais da biblioteca digital: a coleção; os serviços de acesso; o usuário do serviço, no espaço que a biblioteca digital procura organizar”. Logo em seguida, a autora os contextualiza:

- O usuário, entendido como o público em geral ou como usuário individualizado, do qual a biblioteca precisa conhecer as necessidades específicas e as diversas atividades. Devem estar aptos a fazerem uso dos serviços disponíveis;
- Os conteúdos, isto é, os objetos digitais, organizados e estruturados nas coleções digitais segundo normas próprias e distribuídos em rede;
- Os serviços de acesso, caracterizados por interfaces ou serviços mediados pelo pessoal bibliotecário (TAMMARO, 2008, p. 123).

Assim, muito mais do que a plataforma em si, é necessário que existam serviços de qualidade e toda uma gestão para se caracterizar uma biblioteca digital. A necessidade de novos conhecimentos e novas práticas, principalmente em acesso aberto, foram essenciais para o êxito das bibliotecas digitais. Segundo Tammaro (2008), para a comunidade bibliotecária, acrescenta-se dois elementos fundamentais: a biblioteca enquanto instituição e o pessoal bibliotecário especializado para implantar serviços novos ou renovados, sendo que a atenção, nesse caso, recai nos materiais digitais disponibilizados.

Nessa perspectiva, os repositórios digitais<sup>12</sup> das instituições de ensino superior foram implantados no Brasil, principalmente, devido ao trabalho do IBICT, da comunidade bibliotecária e das bibliotecas universitárias, a demanda ampliou-se a partir de uma portaria da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), n. 13 de 2006, bem sucinta que no “[...] art. 1, § 2, relata que devem ser disponibilizados em formato digital teses e dissertações defendidas após março de 2006”. Assim, ocorreu o início das bibliotecas digitais

---

<sup>11</sup> Os eventos internacionais em torno do tema, segundo Tammaro e Sallareli (2008), ocorreram na última década do século XX. E no Brasil teve-se mais projeção no início dos anos 2000.

<sup>12</sup> Antes denominados de Repositórios Institucionais, que obriga o depósito legal de teses e dissertações produzidas dentro das universidades, principalmente as que tiveram apoio/financiamento público.

e a percepção de que era necessário disponibilizar a produção científica da pós-graduação também em formato digital. O IBICT já disseminava o Projeto TEDE no Brasil, porém, com essa portaria, teve-se mais adesão e disseminação dos Repositórios Institucionais (RIs). Muito mais do que um depósito ou um arquivo de itens digitais, os RIs são serviços para atender os usuários sete dias por semana, durante 24 horas. Portanto, os repositórios digitais são uma realidade bem consolidada no país.

## 2.5 REPOSITÓRIO DE DADOS

Na parte tecnológica, a disponibilização dos dados de pesquisa ocorre em plataformas de *softwares* para repositórios de dados. De acordo com Moreno (2018, p. 53),

[...] a criação de infraestrutura e manutenção de repositórios de dados de pesquisa está em curso em diversos países e apresenta-se como um desafio tanto em termos de gestão quanto da representação dos dados ou conjuntos de dados que estão contidos nesses sistemas.

Diversas são as tecnologias disponíveis que permitem a colaboração à distância entre pesquisadores, além da geração, da reunião e do compartilhamento de uma grande quantidade de dados de pesquisa no âmbito da Ciência Aberta (*Open Science*).

Os repositórios de dados são mantidos por conjuntos de ações que viabilizam o armazenamento de dados visando à otimização da recuperação, o que amplia as potencialidades de reuso destes dados entre os pesquisadores. Desta forma, agiliza os processos de investigação e, conseqüentemente, o avanço na ciência. Com uma infraestrutura implementada por repositórios de dados, apoiada por um Plano de Gerenciamento de Dados (PGD) bem fundamentado, os pesquisadores têm aporte propício para depósito de seus conjuntos de dados e busca e recuperação de dados já coletados por outros pesquisadores, que poderão ser reutilizados em suas pesquisas (MONTEIRO, 2017, p. 15).

Os repositórios de dados são bibliotecas digitais, muito semelhantes aos repositórios temáticos e institucionais, porém, alguns mecanismos de armazenamento e de descrição dos metadados são diferentes, permitindo o reuso dos dados e o versionamento destes. Os repositórios de dados se constituem pelos serviços ofertados, a curadoria e a suporte aos pesquisadores antes mesmos da concepção dos dados, a primeira etapa é o planejamento da pesquisa, passa depois para a seleção/definição da amostra, a coleta dos dados, a preparação, a limpeza e a descrição dos metadados, ou seja, toda a curadoria antes de ser depositados em um repositório (SILVA, 2021).

Percebe-se que a infraestrutura tecnológica é importante, porém, deve seguir junto com a gestão, a atribuição de licenças, a descrição dos metadados e a preservação dos dados. Para Sanchez, Vechiato e Vidotti (2019, p. 52), “[...] assim, nascem os Repositórios de Dados, ambientes informacionais digitais que buscam armazenar, organizar, representar, prover acesso, disseminar e preservar dados oriundos de pesquisas científicas”. Existem vários *softwares* para a criação de repositórios de dados, os mais conhecidos e utilizados são o DSPACE, o CKAN e o DATAVERSE.

A criação de ambientes informacionais digitais como o caso dos Repositórios de Dados é de suma importância para o enriquecimento de diversas áreas do conhecimento, pois a cultura de partilha de dados de pesquisa dentro da comunidade potencializa exponencialmente a criação de novas hipóteses de pesquisa, evita a duplicação de esforços e, além disso, traz transparência em relação à utilização dos recursos financeiros que foram investidos nas pesquisas (SANCHEZ; VECHIATO; VIDOTTI, 2019, p. 75).

Segundo Amorim *et al.* (2017, p. 851, tradução nossa),

[...] algumas instituições adotaram repositórios institucionais como base para o depósito de dados, enquanto outras estão experimentando ambientes mais ricos para descrição de dados, apesar da diversidade de fluxos de trabalho existentes.

Sobre os repositórios de dados no Brasil, Paganine e Amaro (2020, p. 187) citam que “[...] as demandas atuais de dados de pesquisa e seu compartilhamento, advêm da ciência aberta e apontam os repositórios de dados, como ferramentas para o armazenamento, organização, compartilhamento e divulgação desses dados”. Porém, esses autores identificaram que são poucos os repositórios instalados, o que, somado à falta de padronização, causa um problema recorrente. O que coloca os repositórios de dados nas contemporâneas práticas de Ciência Aberta.

Se os repositórios desempenham uma função vital na preservação, integridade e divulgação de dados de pesquisa, uma rede de repositórios pode gerar conexões entre as comunidades, aumentando assim a interface entre fontes de dados de diferentes disciplinas e em repositórios específicos ou multidisciplinares (PAVÃO; ROCHA; GABRIEL JÚNIOR, 2018, p. 331).

Segundo Semeler (2017, p. 21), desde 2012, “[...] a distribuição global dos repositórios de dados de pesquisa é catalogada pelo *Registry of Research Data Repositories* (re3data), que é o registro internacional de repositórios de dados de pesquisa”.

O país [Brasil], no entanto, ainda não conta com uma estrutura de apoio à construção de repositórios de dados de pesquisa, nem há um planejamento em nível nacional para a efetivação de uma rede que reúna indivíduos e instituições, paralelamente aos seus conhecimentos e práticas, envolvidos com a promoção de dados abertos da pesquisa brasileira (PAVÃO; ROCHA; GABRIEL JÚNIOR, 2018, p. 331).

Diante desse cenário, considerando o Brasil um país com dimensões continentais, uma instalação centralizada dificultaria a disseminação de conhecimentos, mas uma rede de repositórios e compartilhamento de experiências, promoveria os repositórios, auxiliando até na escolha das plataformas a serem utilizadas. Portanto, uma estrutura que consiga orientar com diretrizes, recursos e financiamento, conforme ocorreu na Europa no projeto FAIRsFAIR (seção 2.8.1), buscando por repositórios digitais confiáveis (certificados), poderá ser um bom exemplo para as iniciativas dos repositórios de dados brasileiros.

## 2.6 A INSTITUCIONALIZAÇÃO DO REPOSITÓRIO DE DADOS

O repositório precisa ser institucionalizado, referendado e reconhecido na instituição. O reconhecimento é importante para que se torne uma exigência ao nível institucional (estratégico), possibilitando, assim, as mudanças práticas na gestão dos dados de pesquisa, ou seja, da cultura organizacional do pesquisador; essa conscientização está na necessidade de adoção de práticas de gestão de dados, como planejar o ciclo de vida dos dados antes da fase de coleta e o seu depósito após o término da pesquisa em repositórios de dados. Segundo Espíndola *et al.* (2018, p. 280),

[...] a governança aplicada a dados científicos está relacionada a planejar diretrizes, criar políticas, determinar papéis e responsabilidades, desenvolver estratégias e processos padronizados, visando que os dados possuam qualidade e estejam disponíveis para quem os necessita.

Políticas são regras mais objetivas, diretas e regulatórias, aplicáveis para ajustar comportamentos, reduzir riscos, definir responsabilidades, etc. Podem advir de princípios. Devem servir de elementos de criação de elos e parcerias entre as partes da empresa envolvidas na GD [Governança de dados]. É importante que as políticas sejam criadas estritamente para alcançar controles objetivos, aplicáveis, usáveis e divulgados, evitando excessos e elementos burocráticos que geram impedimentos e desmotivação. As políticas deverão ser criadas baseadas em processos simples e normalmente se originam de legislações vigentes e diretrizes organizacionais (BARBEIRI, 2019, p. 81).

Como exemplo, a Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA) identifica papéis e responsabilidades na governança de dados, para Bertin *et al.* (2019, p. 199),

[...] a garantia da qualidade e integridade dos dados de pesquisa será de responsabilidade de cada empregado da Empresa, a fim de assegurar autenticidade, integridade, originalidade, rastreabilidade, certificação e reconhecimento de autoria.

De acordo com as autoras, a política determinou ainda: a) constituição e ação do Comitê de Governança de Dados, Informação e Conhecimento, com composição representativa da estrutura organizacional; e b) Comitês Locais de Gestão de Dados, Informação e Conhecimento nas Unidades Descentralizadas da Empresa.

A Política de Governança de Dados, Informação e Conhecimento da Embrapa é um importante instrumento institucional que marca o início de um esforço coletivo interno para alcançar o efetivo gerenciamento dos dados derivados das atividades da Embrapa, especialmente daqueles gerados pela sua área fim, os dados de pesquisa. O reconhecimento dos dados de pesquisa como ativos corporativos evidencia o entendimento de que, uma vez gerenciados e preservados, estes podem ser compartilhados e reutilizados tanto na pesquisa quanto na tomada de decisões e, contribuindo assim para o alcance dos objetivos estratégicos da Empresa (BERTIN *et al.* 2019, p. 201).

A formalização da governança de dados deve partir do nível estratégico, adotando ações, comitês e comissões departamentais para a preparação à implementação de políticas de gestão de dados e os estudos para implantação da tecnologia.

### **2.6.1 Políticas para repositórios de dados**

As políticas para repositórios de dados são diretrizes que trazem estabilidade e definições práticas de como agir e decidir, tanto para os gestores atuais e futuros quanto para os pesquisadores da instituição. Segundo Vieira e Príncipe (2020), no que diz respeito à preparação de políticas de dados de investigação e serviços de apoio para a implementação dos princípios FAIR, muitas das universidades comunicaram a existência de políticas institucionais de dados de investigação e oferta de serviços de apoio para a gestão desses dados.

Uma das dificuldades foi encontrar diretrizes ou políticas de gestão de dados de universidades e instituições brasileiras. A Universidade Federal do Paraná (UFPR) possui diretrizes, porém, o acesso a elas foi negado e solicitado *login* institucional para o acesso. Na Universidade de São Paulo (USP), a resolução n. 7.900, de 11 de dezembro de 2019 e alterada pela Resolução n. 8.094, de 9 de junho de 2021, estabeleceu normas para a “Gestão de Dados

Científicos<sup>13</sup>” na instituição. O conteúdo dessas resoluções traz definições de “dados científicos” e os responsáveis, dentro da universidade, pela política e pelas diretrizes, porém, é um documento genérico e sem a intenção de ser uma política que trata de todos os temas como PGD, licenças abertas, documentação etc. Em sua tese, Almeida (2019) sinteriza os elementos que devem estar presentes nos documentos de políticas de gestão de dados:

Quadro 13 – Elementos de uma política para gestão de dados de pesquisa

<b>Requisitos da política</b>	
Qualidade e padrões de dados	Os investigadores são obrigados a aderir aos padrões internacionais para permitir o acesso e a reutilização.
	A documentação de dados e metadados deve acompanhar os dados para que sejam entendidos por outras pessoas.
Acesso e compartilhamento de dados	Os investigadores são obrigados a disponibilizar dados para serem compartilhados (geralmente mediante a publicação dos resultados ou pouco depois, embora algumas agências permitam períodos de embargo).
	Requisitos para depósito de metadados em catálogo local ou nacional.
Retenção e preservação de dados	Os dados devem ser retidos por um período mínimo determinado.
	Quando possível, os investigadores devem depositar seus dados em arquivo de longo prazo para garantir preservação de seus dados.
Planos de gestão de dados	As propostas de pesquisa devem incluir um Plano de Gestão de Dados.
<b>Disposições comuns às políticas</b>	
Privacidade	Os direitos e a privacidade dos indivíduos que participam da pesquisa devem ser protegidos em todos os momentos. Os dados disponibilizados para uso mais amplo devem estar livres de identificadores que permitam ligações com participantes individuais da pesquisa e variáveis que possam levar à revelação dedutiva da identidade dos indivíduos.
Conhecimento tradicional	No caso de conhecimento local e tradicional, os direitos dos detentores do conhecimento não serão comprometidos.
Dados de natureza sensível	Liberação de dados pode causar danos. Aspectos específicos dos dados podem precisar ser protegidos (por exemplo, locais de ninhos de aves ameaçadas ou locais sagrados).
Propriedade intelectual/Propriedade dos dados	Pode ser necessário, ocasionalmente, atrasar a publicação por curto período, a fim de permitir que os pedidos sejam redigidos.
<b>Outros aspectos</b>	
Princípios	As políticas de dados aderem a um conjunto de princípios abrangentes que articulam seu valor.
Escopo/Cobertura da política	Descreve o escopo dos dados cobertos pela política.
Papéis e responsabilidades	A política identifica as várias partes responsáveis pela gestão de dados nos diferentes estágios do seu ciclo de vida.
Monitoramento e execução	Os meios pelos quais as políticas serão monitoradas ou aplicadas são descritos na política.

Fonte: Shearer (2015, p. 8-9 *apud* ALMEIDA, 2019, p. 85).

<sup>13</sup> Neste trabalho, adotamos o termo “dados de pesquisa”. Porém, a USP adota “dados científicos” para se referir aos dados coletados para subsidiarem as pesquisas acadêmicas da instituição.



## 2.7 SERVIÇO DE SUPORTE À GESTÃO DE DADOS DE PESQUISA NAS UNIVERSIDADES

O serviço de referência é um trabalho de orientação, de acompanhamento, de ajuda, de treinamento e de letramento informacional e digital aos usuários da biblioteca, capacitando-os a utilizarem de recursos tecnológicos para melhorar seu desempenho acadêmico em todos os níveis de formação: graduação, mestrado e doutorado. É um serviço essencial e de alta complexidade pelos conhecimentos que os bibliotecários necessitam para atuarem nesse setor, ensinando, capacitando e treinando os usuários, por exemplo, a utilizarem gerenciadores bibliográficos (Mendeley, Zotero, EndNote), diferentes normas acadêmicas (ABNT, APA, Vancouver, MLA), especificações para busca em diferentes bases de dados do básico ao avançado, normalizações acadêmicas, uso de recursos pela comunidade, como solicitar o ISBN, o DOI, o Programa de Comutação Bibliográfica (COMUT) etc.

O serviço de referência parte dos serviços prestados diretamente ao usuário. Trata-se do processo essencial ao contato entre o usuário e a informação. O começo do serviço de referência é geralmente atribuído a Samuel Sweet Green, que, em 1876, publicou o primeiro artigo sobre a ajuda aos usuários na utilização da biblioteca, sugerindo quatro funções básicas: instruir o usuário sobre como utilizar a biblioteca, responder a suas perguntas, ajudá-lo a selecionar os recursos da biblioteca, promover a biblioteca na comunidade (CUNHA; CAVALCANTI, 2008, p. 334).

Baseado no serviço de referência já prestado nas bibliotecas de acordo com Figueiredo (1992) e nas teses de Almeida (2019) e Semeler (2017) sobre as competências dos bibliotecários de dados e o serviço de suporte a dados de pesquisa, será realizada a proposta de como implantar um setor responsável pela gestão de dados. Assim, um modelo para implantação de repositórios de dados de pesquisa passa necessariamente pelas bibliotecas universitárias que já realizam a gestão dos repositórios digitais, com o depósito das produções intelectuais e da memória científica. Conforme já pensado e estruturado por Almeida (2019) em sua tese, a implantação de um novo setor dentro das bibliotecas denominado “serviço de suporte à gestão de dados de pesquisa”, para a autora, é uma ampliação das tarefas e com bastante similaridade ao serviço de referência/competência em informação.

Para Almeida (2019, p. 18), “[...] as bibliotecas universitárias precisam assumir a responsabilidade de auxiliar na gestão e preservação de conjuntos de dados decorrentes de pesquisas realizadas dentro das universidades”. Em outro exemplo sobre como implantar o acesso aberto nas universidades, Letrouit *et al.* (2021) citam treze recomendações sugeridas

para a França, a cinco e a oito dizem respeito à implantação de serviços de gestão de dados pelas bibliotecas universitárias:

**Recomendação n.º 5:** desenvolver uma estratégia da instituição para a gestão de dados de investigação assente numa oferta de serviços básicos que consiste na sensibilização e formação de doutorandos e investigadores em gestão de dados, independentemente da sua ancoragem disciplinar. Incluir um aumento gradual de recursos e serviços de acordo com os objetivos sucessivos estabelecidos pelo estabelecimento. [...] **Recomendação n.º 8:** incentivar pesquisadores e bibliotecários a trabalharem juntos assim que surgir um projeto de pesquisa para se conhecerem melhor e melhorarem a eficiência, principalmente na questão da gestão de dados (LETROUIT *et al.*, 2021, p. 49 -51, tradução nossa, grifo nosso).

Segundo Silva (2019), para a gestão de dados, são necessários seis grandes núcleos: agências de financiamento, plano de gestão, ciclo de vida, repositórios, papel dos bibliotecários e papel dos próprios pesquisadores. As agências de financiamento são, geralmente, os órgãos que primeiro realizam a cobrança da gestão de dados em seus editais, fazendo com que as universidades e seus pesquisadores realizem as adequações necessárias para o cumprimento dessas práticas.

## 2.8 CERTIFICAÇÃO CORETRUSTSEAL

As certificações existentes para a auditoria e para a certificação de repositórios digitais confiáveis são definidas em três níveis: básico, estendido ou formal. O processo vai de autoavaliação, auditoria até a certificação.

- I. CoreTrustSeal é um novo selo (tipo de reconhecimento) para substituir e unificar a certificação *Data Seal of Approval (DSA)*, da *Data Archiving and Networked Services (DANS)*, da Holanda, e a Certificação *World Data System (WDS)*, do *International Science Council (ICS)*. “A Core Trust Seal é uma organização internacional, comunitária, não governamental e sem fins lucrativos que promove infraestruturas de dados sustentáveis e confiáveis” (CoreTrustSeal, 2022). Possui 16 requisitos a serem avaliados.
- II. A certificação DIN 31644 (2012) é da *Network of Expertise in Long-Term Storage of Digital Resources (NESTOR)*, da Alemanha. Segundo o *Digital Preservation Handbook* (2015, tradução nossa), o processo de certificação estendida realizado pela Nestor possui mais de 30 requisitos divididos em três partes: estrutura organizacional,

gerenciamento de objetos e infraestrutura e segurança. A obtenção do certificado leva em torno de três meses.

- III. A Certificação ISO 16363 (2012), *Audit and certification of trustworthy digital repositories* (Auditoria e certificação de repositórios digitais confiáveis), é uma certificação formal com validade de cinco anos, realizada por auditores externos (certificados pela ISO 16919), o repositório precisa possuir alguma outra certificação, CoreTrustSeal ou DIN.

Um repositório certificado é confiável, garantindo a qualidade da preservação de seus objetos digitais. A certificação é um amadurecimento dos serviços e reconhecimento de sua qualidade e responsabilidade, principalmente, uma garantia para os pesquisadores. O apoio para as certificações no contexto europeu surgiu para reconhecer que os repositórios possuem boas práticas, estão formalmente apoiados por suas instituições/organizações e para que seus pesquisadores tenham o hábito de compartilhar seus dados e também reusar dados de terceiros.

Atualmente, os padrões de certificação estão disponíveis em diferentes níveis, desde o nível básico até os níveis estendidos e formais. Mesmo no nível básico, a certificação oferece muitos benefícios para um repositório e suas partes interessadas. A certificação central envolve um processo minimamente intensivo pelo qual os repositórios de dados fornecem evidências de que são sustentáveis e confiáveis. Um repositório primeiro conduz uma autoavaliação interna, que é então revisada por colegas da comunidade. Essas avaliações ajudam as comunidades de dados – produtores, repositórios e consumidores – a melhorar a qualidade e a transparência de seus processos e aumentar a conscientização e a conformidade com os padrões estabelecidos. Essa abordagem da comunidade garante uma atmosfera inclusiva na qual o repositório candidato e os revisores interagem estreitamente (CORE TRUST SEAL, 2022, tradução nossa).

A CoreTrustSeal é a certificação que interessa a este trabalho, pois, os repositórios de dados brasileiros (que ainda são poucos) podem auxiliar no planejamento destes. E os que estão em fase de concepção podem seguir os critérios da certificação, já pensando num serviço de excelência e de qualidade. Um repositório certificado garante que ele realiza boas práticas, possui qualidade, assegurando que os dados estão seguros e serão preservados. É uma forma de demonstrar aos pesquisadores que podem confiar naquele repositório, afinal, seus dados estarão resguardados e íntegros. A certificação CoreTrustSeal para repositórios de dados, segundo relatório<sup>14</sup> da própria organização, é necessária para que os financiadores nacionais e

---

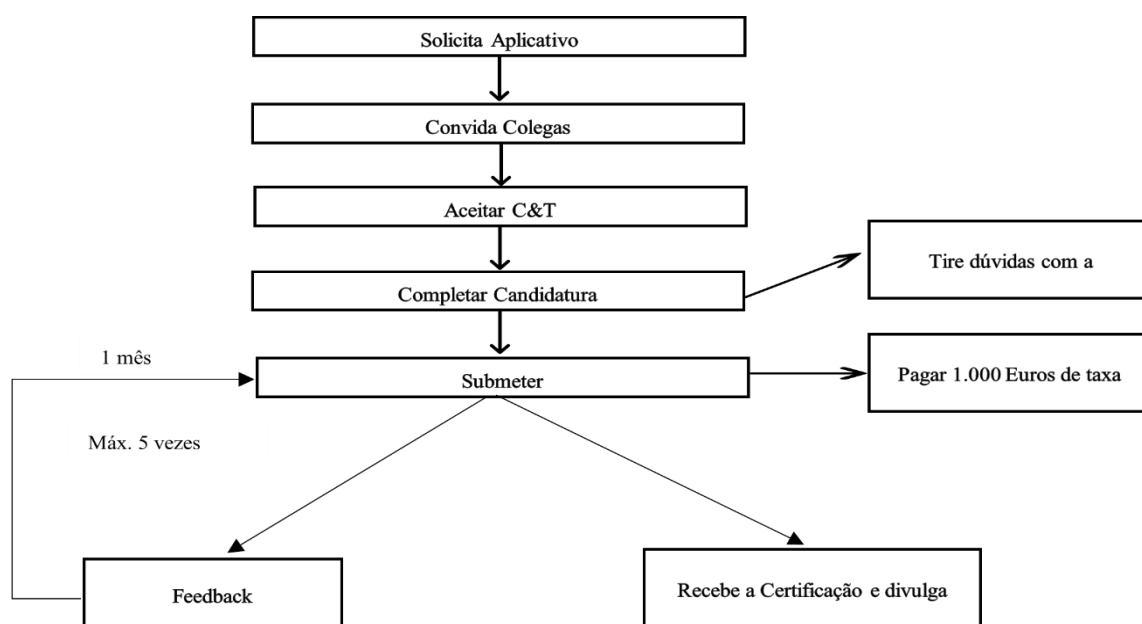
<sup>14</sup> CORETRUSTSEAL STANDARDS AND CERTIFICATION BOARD. CoreTrustSeal Trustworthy Data Repositories Requirements 2020–2022. [S.l.]: Zenodo, 20 nov. 2019. Disponível em: <https://doi.org/10.5281/zenodo.3638211>.

internacionais exigem políticas de gerenciamento de dados, que podem armazenamento e acessibilidade a longo prazo.

A certificação pode ser uma importante contribuição para garantir a confiabilidade e durabilidade dos repositórios de dados, portanto, o potencial para compartilhar dados por um longo período. Ao se tornarem certificados, os repositórios podem demonstrar a seus usuários e financiadores que uma autoridade independente os avaliou e endossou sua confiabilidade (CORETRUSTSEAL, 2022, não paginado).

Para obter a certificação é necessário submeter documentos e pagar uma taxa administrativa de mil euros<sup>15</sup>, os consórcios de repositórios com dez ou mais podem obter descontos de 25% no valor de cada assinatura. A certificação possui validade de três anos, sendo necessário o pagamento de nova taxa e o envio de toda a documentação atualizada para recertificação.

Figura 9 – Fluxo da certificação CoreTrustSeal



Fonte: Adaptada e traduzida de Edmunds (2018).

De acordo com Rorie Edmunds (membro do conselho CoreTrustSeal *ex officio*, World Data System), no Webinar<sup>16</sup> de 30 de outubro de 2018, cada um dos dezesseis critérios avaliados pela CoreTrustSeal precisa de conceito três ou quatro. Os conceitos um e dois servem

<sup>15</sup> O mesmo valor é cobrado desde 2018, ano em que o certificado foi implementado.

<sup>16</sup> I Webinars sobre a certificação CoreTrustSeal. Disponível em:

<https://www.youtube.com/watch?v=gs9HTzxbx0s>. Acesso em: 18 ago. 2022.

II coretrustseal certification. Disponível em: <https://www.coretrustseal.org/why-certification/requirements/>. Acesso: 18 ago. 2022.

somente como parâmetro para autoavaliação interna dos repositórios de dados, que antes de submeterem para a certificação, devem buscar a implementação de todos os critérios, estes são independentes um critério do outro, assim, todos precisam de conceito mínimo de três, não sendo um somatório de média ponderada.

Quadro 14 – Nível de conformidade para cada um dos requisitos

Conceito	Critérios
0 (zero)	Não aplicável.
1 (um)	O repositório ainda não considerou isso.
2 (dois)	O repositório tem conceito teórico.
3 (três)	O repositório está em fase de implementação.
4 (quatro)	A diretriz foi totalmente implementada no repositório.

Fonte: Adaptado e traduzido de Edmunds (2018).

Quadro 15 – Critérios CoreTrustSeal

	Nº	DESCRIÇÃO DO REQUISITO	OBSERVAÇÃO
<b>0. Contexto e informações básicas</b>	<b>R0</b>	Forneça contexto para o seu repositório.	-
<b>INFRAESTRUTURA ORGANIZACIONAL</b>			
<b>1. Missão/Escopo</b>	<b>R1</b>	O repositório tem a missão explícita de fornecer acesso e preservar os dados em seu domínio.	-
<b>2. Licenças</b>	<b>R2</b>	O repositório mantém todas as licenças aplicáveis que abrangem o acesso e o uso de dados e monitora a conformidade.	As disposições éticas e de privacidade que afetam as licenças são tratadas em R4. A garantia de que as licenças de depósito fornecem direitos suficientes para o repositório manter, preservar e oferecer acesso aos dados deve ser coberta pelo R10.
<b>3. Continuidade de acesso</b>	<b>R3</b>	O repositório tem um plano de continuidade para garantir o acesso contínuo e a preservação do seu acervo.	Requisito relacionado à governança. Os aspectos técnicos de continuidade de negócios e de planejamento de desastres e de sucessão devem ser abordados em R15.
<b>4. Confidencialidade/ética</b>	<b>R4</b>	O repositório garante, na medida do possível, que os dados sejam criados, curados, acessados e usados em conformidade com as normas disciplinares e éticas.	Detalhes sobre quaisquer licenças em alinhamento com tais disposições éticas e de privacidade devem ser cobertos em R2.
<b>5. Infraestrutura organizacional</b>	<b>R5</b>	O repositório tem financiamento adequado e um número suficiente de pessoal qualificado, gerido através de um sistema claro de governança para levar a cabo a missão de forma eficaz.	O acesso a aconselhamento especializado objetivo, além do fornecido por pessoal qualificado, é coberto em R6.
<b>6. Orientação especializada</b>	<b>R6</b>	O repositório adota mecanismo(s) para garantir orientações e <i>feedbacks</i> contínuos de especialistas (internos ou externos,	Este requisito procura confirmar que o repositório tem acesso a aconselhamento especializado objetivo, além do fornecido pelo

		incluindo orientação científica, se relevante).	pessoal qualificado mencionado em R5.
<b>GERENCIAMENTO DE OBJETOS DIGITAIS</b>			
<b>7. Integridade e autenticidade dos dados</b>	<b>R7</b>	O repositório garante a integridade e a autenticidade dos dados.	-
<b>8. Avaliação</b>	<b>R8</b>	O repositório aceita dados e metadados com base em critérios definidos para garantir relevância e compreensão para os usuários dos dados.	Este requisito abrange os critérios de seleção aplicados no ponto de depósito. A qualidade dos dados e a melhoria durante o processo de curadoria devem ser cobertas pelo R11.
<b>9. Procedimentos de armazenamento documentados</b>	<b>R9</b>	O repositório aplica processos e procedimentos documentados no gerenciamento do armazenamento de dados em arquivo.	Detalhes sobre a implementação técnica de armazenamento devem ser cobertos em R15 e arranjos específicos para segurança física e lógica em R16.
<b>10. Plano de preservação</b>	<b>R10</b>	O repositório assume a responsabilidade pela preservação a longo prazo e gere essa função de forma planejada e documentada.	Os direitos relativos ao acesso e ao uso de dados e o monitoramento de sua conformidade devem ser cobertos pelo R2.
<b>11. Qualidade dos dados</b>	<b>R11</b>	O repositório possui conhecimento adequado para tratar da qualidade de dados técnicos e metadados e garante que informações suficientes estejam disponíveis para que os usuários finais façam avaliações relacionadas à qualidade.	Este requisito refere-se a padrões de qualidade de dados e garantia durante a curadoria. Os critérios de seleção são abordados em R8.
<b>12. Fluxos de trabalho</b>	<b>R12</b>	O arquivamento ocorre de acordo com fluxos de trabalho definidos, desde a ingestão até a disseminação.	Este requisito confirma que todos os fluxos de trabalho estão documentados.
<b>13. Descoberta e identificação de dados</b>	<b>R13</b>	O repositório permite que os usuários descubram os dados e os consultem de forma persistente por meio da citação adequada.	-
<b>14. Reutilização de dados</b>	<b>R14</b>	O repositório permite a reutilização dos dados ao longo do tempo, garantindo que os metadados apropriados estejam disponíveis para apoiar a compreensão e o uso dos dados.	-
<b>TECNOLOGIA</b>			
<b>15. Infraestrutura técnica</b>	<b>R15</b>	O repositório funciona em sistemas operacionais bem suportados, em outro <i>software</i> de infraestrutura central e está usando tecnologias de <i>hardware</i> e <i>software</i> apropriadas aos serviços que fornece à sua comunidade designada.	Os aspectos de governança de continuidade de negócios, de planejamento de desastres e de planejamento de sucessão devem ser abordados em R3. Os detalhes sobre o processo de armazenamento devem ser abordados em R9. Os arranjos de segurança são cobertos em R16.
<b>16. Segurança</b>	<b>R16</b>	A infraestrutura técnica do repositório fornece proteção da instalação e de seus dados, produtos, serviços e usuários.	Os processos de armazenamento e de infraestrutura técnica que utilizam essas medidas de segurança devem ser abordados em R9 e R15, respectivamente.

Fonte: Adaptado e traduzido de *Core Trustworthy Data Repositories Requirements 2020-2022*<sup>17</sup>.

Figura 10 – Geolocalização dos repositórios certificados com a CoreTrustSeal ou anteriores



Fonte: Captura de tela realizada pelo autor<sup>18</sup> (2022).

Apresenta-se a lista dos repositórios certificados com a CoreTrustSeal ou anteriores (WDS ou DAS). A América Latina não possui nenhum repositório certificado até o momento [maio de 2022], dessa forma, é importante que comece a integração do continente na implementação de repositório de dados confiáveis.

### 2.8.1 O projeto FAIRsFAIR: em busca da certificação Core Trust Seal

O *Fostering Fair Data Practices in Europe* (FAIRsFAIR) foi um projeto iniciado em 2019, ele “[...] visa fornecer soluções práticas para o uso dos princípios de dados FAIR ao longo do ciclo de vida dos dados de pesquisa” (FAIRsFAIR, 2022), tem a duração de 36 meses, com início em março de 2019 e orçamento de dez milhões de euros<sup>19</sup>. É financiado pelo Projeto Horizon 2020 e contribui para o *European Open Science Cloud* (EOSC) em busca de uma

<sup>17</sup> CORETRUSTSEAL STANDARDS AND CERTIFICATION BOARD. CoreTrustSeal Trustworthy Data Repositories Requirements 2020–2022. 20 nov. 2019. Disponível em: <https://doi.org/10.5281/zenodo.3638211>. Acesso em: 19 ago. 2022.

<sup>18</sup> CORE CERTIFIED REPOSITORIES. Disponível em: <https://www.coretrustseal.org/why-certification/certified-repositories/>. Acesso: 08 maio 2022.

<sup>19</sup> DILLO, Ingrid. FAIRsFAIR 2021 - FAIRsFAIR in the context of the EOSC. Disponível em: <https://www.youtube.com/watch?v=lpjw6vHCgSM&list=PLXXwep9JCnSQzgJfoS4A-zkt7IEdhnapQ>. Acesso: 08 maio 2022.

infraestrutura funcional de dados FAIR, pensando no desenvolvimento de padrões globais para a certificação FAIR de repositórios. O projeto tem dupla abordagem para desenvolver um padrão para certificação: (1) suporte para repositórios de dados para obter a certificação CoreTrustSeal; (2) suporte para a “FAIRification” de repositórios e melhoria da FAIRness e interoperabilidade dos dados. Segundo Vieira e Príncipe (2020, p. 2),

[...] no âmbito das recomendações para os repositórios de dados de investigação, o projeto FAIRsFAIR contribui para a identificação e definição de políticas e práticas para uma adoção ampla dos princípios FAIR e no desenvolvimento de normas para a certificação FAIR de repositórios.

Conforme esses autores, o projeto possui um conjunto de orientações para os repositórios acolherem objetos digitais FAIR e para que esses sejam verdadeiramente FAIR, os requisitos devem ser ao nível técnico e organizacional.

FAIRsFAIR visa fornecer soluções para o uso dos princípios de dados FAIR ao longo do ciclo de vida dos dados de pesquisa. Garantir que recursos de pesquisa como dados, software e serviços possam ser prestados e mantidos, FAIR é a principal contribuição deste projeto para o EOSC. A FAIRsFAIR estabeleceu a Força de Sincronização para maximizar a eficiência de custos e eliminar a duplicação de esforços, alinhando e cooperando com iniciativas que apoiam a construção do EOSC (EOSCSECRETARIAT.EU, 2020, não paginado, tradução nossa).

Com uma chamada aberta e um incentivo financeiro, o projeto selecionou dez repositórios europeus para atenderem totalmente aos princípios FAIR e obterem a certificação CoreTrustSeal e outros doze repositórios para a melhoria de interoperabilidade, concorreram, ao todo, 73 repositórios. Segundo o relatório, “os gestores e representantes dos repositórios foram convidados a candidatar-se a receber um incentivo financeiro (€ 1.000,00), bem como apoio dedicado à pré-submissão ao processo de certificação CoreTrustSeal”.

Dez repositórios europeus foram apoiados em sua jornada para se tornarem mais confiáveis e habilitados para FAIR. O programa consistiu em várias formas de suporte, incluindo workshops colaborativos e suporte da comunidade de pares, bem como chamadas de suporte individuais e ferramentas práticas, tudo para ajudar a preparar os repositórios selecionados para a certificação CoreTrustSeal. Em um processo iterativo, os repositórios elaboraram autoavaliações e receberam feedback de especialistas da equipe de suporte FAIRsFAIR, na preparação da submissão formal do CoreTrustSeal (VERBURG *et al.*, 2021, p. 4, tradução nossa).

Segundo Huigen e Von Stein (2021), os dez repositórios entregaram rascunhos que foram revisados por cinco equipes de dois revisores cada, em agosto e setembro de 2020, esses fizeram avaliações por pares de teste de trabalhos enviados. O projeto usou o Google Drive



como ferramenta para compartilhar a documentação e também criou um grupo no Google, chamado Fórum de Certificação FAIRsFAIR, onde repositórios e revisores do projeto puderam trocar perguntas e ideias.

O projeto FAIRsFAIR teve sete grupos de trabalho: (1) Gestão de Projetos e Sustentabilidade; (2) Práticas FAIR: semântica, interoperabilidade e serviços; (3) Política e Prática FAIR; (4) Certificação FAIR; (5) Engajamento, comunicação e aceitação; (6) Centro de Competência FAIR; e (7) FAIR *Data Science* e profissionalização. O projeto capacitou os participantes/concorrentes com: oficinas colaborativas; teste com a revisão por pares dos aplicativos CoreTrustSeal por especialistas, como a preparação para submissão formal ao Processo CoreTrustSeal; apoio e aconselhamento individualizado; uma variedade de ferramentas para apoiar serviços e indivíduos em suas práticas de dados FAIR; um *helpdesk* dedicado e apoiado à comunidade de pares.

[...] o FAIRsFAIR ofereceu uma série de nove webinars que visavam ajudar uma comunidade muito mais ampla de gerentes de repositório a se familiarizar com as práticas de habilitação do FAIR. Cada webinar de uma hora forneceu uma visão geral de uma atividade específica de habilitação do FAIR, informações compartilhadas sobre desenvolvimentos recentes no FAIRsFAIR e outras iniciativas, além de oferecer exemplos de boas práticas, dicas práticas e recomendações<sup>20</sup> (SUPPORT PROGRAMME FOR DATA REPOSITORIES, [2022], não paginado).

Segundo Verburg *et al.* (2021, tradução nossa), um dos sucessos das oficinas realizadas foi a presença de 40 revisores da certificação CoreTrustSeal, de um total de 57 revisores. O projeto está documentado por relatórios<sup>21</sup> que abordam questões como: plano de gerenciamento de dados; plano de sustentabilidade; infraestrutura de pesquisa; princípios de dados FAIR em todo o ciclo de vida dos dados de pesquisa; cultura de dados FAIR; políticas e práticas de dados FAIR; modelo de maturidade para dados FAIR em repositórios FAIR; competências FAIR e outros inúmeros assuntos.

O projeto auxiliou na busca por qualidade, padrões e boas práticas, ou seja, a maturidade no trabalho com os dados de pesquisa. A certificação CoreTrustSeal é uma garantia de segurança, perenidade e preservação dos dados. É uma forma de mostrar que o repositório possui critérios sólidos, políticas e boas práticas na gestão dos dados nele depositados. Ao final o projeto, “FAIRsFAIR fornecerá uma plataforma para usar e implementar os princípios FAIR

---




<sup>20</sup> SUPPORT PROGRAMME FOR DATA REPOSITORIES. Disponível em:

<https://www.fairsfair.eu/application-results-open-call-data-repositories>. Acesso em: 19 ago. 2022.

<sup>21</sup> FAIRsFAIR. Deliverables & Milestones. Disponível em: <https://www.fairsfair.eu/reports-deliverables>. Acesso em: 19 ago. 2022.

no trabalho diário dos provedores e repositórios de dados de pesquisa europeus” (FAIRsFAIR, 2022). Assim, o projeto desenvolveu algumas ferramentas que facilitam o trabalho dos gestores e pesquisadores para obterem dados e repositórios FAIR:

Quadro 16 – *Software* e ferramentas desenvolvidas pelo FAIRsFAIR

 <p>www.fairsfair.eu/fair-aware</p>	<p>FAIR-Aware é uma ferramenta online que ajuda pesquisadores e gerentes de dados a avaliar o quanto eles sabem sobre os requisitos para tornar os conjuntos de dados localizáveis, acessíveis, interoperáveis e reutilizáveis (FAIR) antes de carregá-los em um repositório de dados.</p>
 <p><a href="https://www.fairsfair.eu/fairsfair-data-object-assessment-metrics-request-comments">https://www.fairsfair.eu/fairsfair-data-object-assessment-metrics-request-comments</a></p>	<p>17 métricas mínimas viáveis propostas pela FAIRsFAIR para a avaliação sistemática de objetos de dados FAIR.</p>
 <p><a href="https://fairsfair.eu/f-uji-automated-fair-data-assessment-tool">https://fairsfair.eu/f-uji-automated-fair-data-assessment-tool</a></p>	<p>O F-UJI é um serviço baseado em REST, e está conduzindo uma avaliação programática da FAIRness de conjuntos de dados de pesquisa. A avaliação F-UJI é baseada em 16 das 17 principais métricas de avaliação de objetos FAIR desenvolvidas dentro do FAIRsFAIR e cada uma correspondendo a uma parte ou à totalidade de um princípio FAIR.</p>

Fonte: Adaptado de FAIRsFAIR (2022) e Vieira e Príncipe (2020).

Portanto, o FAIRsFAIR contribui com os conhecimentos práticos e com os relatórios publicados para o desenvolvimento de infraestrutura global para dados FAIR, com recomendações, ferramentas/aplicativos que poderão ser usados/aplicados também em repositórios brasileiros. Porém, a realidade brasileira é diferente, desde os incentivos financeiros até a falta de uma rede coordenada para implantação de repositórios de dados, conforme considerado por Rocha *et al.* (2021); entretanto, com os exemplos do projeto FAIRsFAIR e com todas as ferramentas mencionadas (Quadro 16), pode-se buscar a maturidade nos repositórios de dados que estão sendo instalados no país.

### 3 PROCEDIMENTOS METODOLÓGICOS

Neste capítulo, apresenta-se a metodologia utilizada durante o percurso desta pesquisa. Segundo Cunha (1982) e Marconi e Lakatos (2010), a metodologia de pesquisa deve ser escolhida de acordo com o tipo de pesquisa a ser desenvolvida.

#### 3.1 CARACTERIZAÇÃO DA PESQUISA

Quanto aos objetivos, esta pesquisa é classificada como exploratória e descritiva, com abordagem qualitativa, pois, busca conhecer, entender e aprofundar o conhecimento sobre a implantação dos repositórios de dados, de acordo com o objetivo geral: “Propor um modelo para a implantação da gestão de dados de pesquisa nas universidades brasileiras”. Segundo Cerro e Bervian (2002), os estudos exploratórios têm por objetivo familiarizar-se com o fenômeno ou obter nova percepção sobre ele e descobrir novas ideias. A pesquisa exploratória realiza descrições precisas e quer descobrir as relações existentes entre os elementos componentes de uma situação.

**Pesquisa exploratória:** quando a pesquisa se encontra na fase preliminar, tem como finalidade proporcionar mais informações sobre o assunto que vamos investigar, possibilitando sua definição e seu delineamento, isto é, facilitar a delimitação do tema da pesquisa; orientar a fixação dos objetivos e a formulação das hipóteses ou descobrir um novo tipo de enfoque para o assunto (PRODANOV; FREITAS, 2017, p. 51-52, grifo do autor).

A pesquisa descritiva, de acordo com Cerro e Bervian (2002), observa, registra, analisa e correlaciona fatos ou fenômenos (variáveis) sem manipulá-los. Para os autores, a pesquisa descritiva, em suas diversas formas, trabalha sobre dados ou fatos colhidos da própria realidade.

#### 3.2 QUANTO AOS PROCEDIMENTOS

Os procedimentos serão classificados de acordo com cada objetivo específico. Portanto, os procedimentos classificam-se como: bibliográfico, documental e experimental.

##### 3.2.1 Pesquisa bibliográfica

A pesquisa bibliográfica foi utilizada neste trabalho por se tratar de um tema em ascensão para os bibliotecários brasileiros, sendo assim, necessário a construção de conhecimentos acerca do objeto de estudo, utilizando materiais bibliográficos publicados,

[...] constituído[s] principalmente de: livros, revistas, publicações em periódicos e artigos científicos, jornais, boletins, monografias, dissertações, teses, material cartográfico, internet, com o objetivo de colocar o pesquisador em contato direto com todo material já escrito sobre o assunto da pesquisa” (PRODANOV; FREITAS, 2017, p. 54).

Atendendo, assim, o objetivo específico de “Levantar os fundamentos teóricos e recomendações relacionadas à gestão de dados e à implantação de repositórios de dados de pesquisa;” (PRODANOV; FREITAS, 2017, p. 54).

Para Cervo e Bervian (2002), qualquer espécie de pesquisa, em qualquer área, supõe e exige uma pesquisa bibliográfica prévia, quer para o levantamento do estado da arte do tema, quer para a fundamentação teórica ou, ainda, para justificar os limites e as contribuições da própria pesquisa. A construção da ciência se faz de maneira lenta e progressiva, tendo como base os estudos anteriormente publicados pela comunidade científica, desse modo, nunca se parte do zero para criar algo, por isso a importância da pesquisa bibliográfica.

A pesquisa bibliográfica procura explicar um problema a partir de referências teóricas publicadas em documentos. Pode ser realizada independentemente ou como parte da pesquisa descritiva ou experimental. Em ambos os casos, busca conhecer e analisar as contribuições culturais ou científicas do passado existentes sobre um determinado assunto, tema ou problema (CERVO; BERVIAN, 2002, p. 65).

A definição de termos para o início das buscas bibliográficas começou a partir da leitura de um artigo sobre o tema e a análise das suas referências, visando chegar a outros termos para uma busca com relevância. Após essa leitura, confeccionou-se alguns termos, listados no quadro abaixo, que têm relação com os repositórios de dados, permitiu-se, assim, identificar o assunto principal e os secundários<sup>22</sup> relacionados com os repositórios de dados.

Quadro 17 – Lista de termos para pesquisa em bases de dados

<b>Português</b>	<b>Inglês</b>	<b>Espanhol</b>
Repositório de dados	<i>Data repositior</i>	<i>Repositorios de datos</i>
Dados de pesquisa	<i>Research data</i>	<i>Datos de investigación</i>
Ciência aberta	<i>Open Science</i>	<i>Ciencia abierta</i>
Curadoria digital	<i>Data curation</i>	<i>Curaduría de datos</i>
Gestão de dados de pesquisa	<i>Data Management</i>	<i>Gestión de datos de investigación</i>
Propriedade intelectual	<i>Intellectual property</i>	<i>Propiedad intelectual</i>
Plano de gerenciamento de dados	<i>Data manegement plan</i>	<i>Plan de gestión de datos</i>
Licenças de uso	<i>Use license</i>	<i>Licencia de uso</i>
Propriedade intelectual	<i>Intellectual property</i>	<i>Propiedad intelectual</i>
Direito autoral	<i>Copyright</i>	<i>Derecho de autor</i>

<sup>22</sup> Assunto principal refere-se a experiências de implantação de *software* e de customizações. Assunto secundário é considerado o tema que não tem relação direta com o objetivo principal da pesquisa do mestrado, mas é importante para se construir e gerenciar um repositório de dados.

Dataverse	<i>Dataverse</i>	<i>Dataverse</i>
-----------	------------------	------------------

Fonte: Elaborado pelo autor (2021).

A pesquisa bibliográfica foi realizada em quatro bases de periódicos e/ou resumos de periódicos e uma base de teses e dissertações, sendo: a) Base de Dados Referenciais de Artigos de Periódicos em Ciência da Informação (BRAPCI); b) Scientific Electronic Library Online (SciELO); c) Biblioteca Digital de Teses e Dissertações (BDTD); d) Library and Information Science Abstracts (LISA)<sup>23</sup>; e) Portal de Periódicos Capes. As referências citadas nos materiais bibliográficos, como artigos, teses, dissertações e capítulos de livros, foram utilizadas como fontes para novas bibliografias. Assim, os materiais bibliográficos utilizados não correspondem somente aos itens recuperados em bases de dados.

### 3.2.2 Pesquisa documental

A pesquisa documental se diferencia da pesquisa bibliográfica somente pelo tipo de fonte e pelo tratamento oferecido ao documento, pois este não passa pelo crivo de especialistas, ou seja, a revisão por pares. Segundo Prodanov e Freitas (2017, p. 55-56), “a utilização da pesquisa documental é destacada no momento em que podemos organizar informações que se encontram dispersas, conferindo-lhe uma nova importância como fonte de consulta”. Na pesquisa documental, as fontes são, geralmente, documentos arquivísticos, relatórios, manuais de como fazer e demais tipos de documentos não publicados com a finalidade acadêmica, mas que podem possuir valor histórico e/ou dados relevantes para embasar estudos científicos, como exemplo, temos o Censo Demográfico do Instituto Brasileiro de Geografia e Estatística (IBGE).

Nesta pesquisa, as fontes documentais utilizadas são documentos publicados em sites e eventos acadêmicos em formato de vídeo. Portanto, pode-se defini-las como documentos digitais, porque são fontes de informações não revisadas por pares, manuais, guias e demais publicações dispersas, porém, são essenciais para o aprendizado sobre a gestão e a curadoria de dados e de repositórios de dados.

---

<sup>23</sup> Vale ressaltar que a base de dados LISA, na época do levantamento, em janeiro de 2021, ainda integrava o Portal de Periódicos da Capes, via acesso CAFe. A base, ainda durante a pesquisa, deixou o Portal de Periódicos, porém, como já havíamos realizado a coleta e estávamos na fase de limpeza dos dados, a base permaneceu neste estudo.

Quadro 18 – Fontes documentais utilizadas

Nome das fontes/Identificação	Endereço eletrônico
CICLO DE VIDA DATAONE	<a href="https://old.dataone.org/data-life-cycle">https://old.dataone.org/data-life-cycle</a>
CORE TRUST SEAL	<a href="https://www.coretrustseal.org/about/">https://www.coretrustseal.org/about/</a>
DATA CITE	<a href="https://datacite.org/value.html">https://datacite.org/value.html</a>
DATA DOCUMENTATION INITIATIVE (DDI)	<a href="http://www.ddialliance.org/">http://www.ddialliance.org/</a>
FAIRSF AIR	<a href="https://www.fairsfair.eu/the-project">https://www.fairsfair.eu/the-project</a> <a href="https://www.fairsfair.eu/reports-deliverables">https://www.fairsfair.eu/reports-deliverables</a>
FERRAMENTA FAIR AWARE	<a href="http://www.fairsfair.eu/fair-aware">www.fairsfair.eu/fair-aware</a>
FERRAMENTA DATA OBJECT ASSESSMENT METRICS	<a href="https://www.fairsfair.eu/fairsfair-data-object-assessment-metrics-request-comments">https://www.fairsfair.eu/fairsfair-data-object-assessment-metrics-request-comments</a>
FERRAMENTA AUTOMATED FAIR DATA ASSESSMENT TOOL (F-UJI)	<a href="https://fairsfair.eu/f-uji-automated-fair-data-assessment-tool">https://fairsfair.eu/f-uji-automated-fair-data-assessment-tool</a>
FAIRSF AIR 2021 Public Event	<a href="https://www.fairsfair.eu/events/fairsfair-2021-public-workshop">https://www.fairsfair.eu/events/fairsfair-2021-public-workshop</a>
FAPESP	<a href="https://www.fapesp.br/openscience/">https://www.fapesp.br/openscience/</a>
FERRAMENTA DMP ONLINE	<a href="https://dmponline.dcc.ac.uk/">https://dmponline.dcc.ac.uk/</a> <a href="https://github.com/DigitalCurationCentre/roadmap/wiki/Help-for-administrators">https://github.com/DigitalCurationCentre/roadmap/wiki/Help-for-administrators</a> <a href="https://dmponline.dcc.ac.uk/help">https://dmponline.dcc.ac.uk/help</a>
FERRAMENTA DMP TOOL	<a href="https://dmptool.org/">https://dmptool.org/</a>
FORCE 11	<a href="https://force11.org/info/about-force11/">https://force11.org/info/about-force11/</a>
GESTÃO DE DADOS USP	<a href="http://www.sibi.usp.br/?p=6189">http://www.sibi.usp.br/?p=6189</a>
GO FAIR	<a href="https://www.go-fair.org/">https://www.go-fair.org/</a> <a href="https://www.go-fair.org/go-fair-initiative/">https://www.go-fair.org/go-fair-initiative/</a>
GO FAIR BRASIL	<a href="https://www.go-fair-brasil.org/">https://www.go-fair-brasil.org/</a>
GUIA DO USUÁRIO DATAVERSE	<a href="https://guides.dataverse.org/en/5.10.1/user/">https://guides.dataverse.org/en/5.10.1/user/</a>
HANDLE NET	<a href="http://www.handle.net/">http://www.handle.net/</a>
LICENÇAS CREATIVE COMMONS	<a href="https://creativecommons.org/licenses/?lang=pt">https://creativecommons.org/licenses/?lang=pt</a>
OPEN DATA COMMONS	<a href="https://opendefinition.org/od/2.1/en/">https://opendefinition.org/od/2.1/en/</a> <a href="https://opendatacommons.org/faq/licenses/">https://opendatacommons.org/faq/licenses/</a> <a href="https://opendatacommons.org/licenses/pddl/1-0/">https://opendatacommons.org/licenses/pddl/1-0/</a> <a href="https://opendatacommons.org/licenses/odbl/1-0/">https://opendatacommons.org/licenses/odbl/1-0/</a> <a href="https://opendatacommons.org/licenses/by/1-0/">https://opendatacommons.org/licenses/by/1-0/</a>
RE3DATA	<a href="https://www.re3data.org/">https://www.re3data.org/</a> <a href="https://doi.org/10.17616/R3D">https://doi.org/10.17616/R3D</a>
RELATÓRIOS DA REDE DE DADOS DE PESQUISA – RDP/UFRGS	<a href="https://cedap.ufrgs.br/jspui/handle/20.500.11959/1199">https://cedap.ufrgs.br/jspui/handle/20.500.11959/1199</a>
RESEARCH DATA LIFECYCLE - UK DATA ARCHIVE	<a href="https://ukdataservice.ac.uk/learning-hub/research-data-management/">https://ukdataservice.ac.uk/learning-hub/research-data-management/</a>
SOFTWARE DATAVERSE	<a href="https://dataverse.org/">https://dataverse.org/</a>
SUPORTE E-CIÊNCIA RNP	<a href="https://ajuda.rnp.br/sec/ ciencia/dadosabertos/gt-rdp-brasil">https://ajuda.rnp.br/sec/ ciencia/dadosabertos/gt-rdp-brasil</a>
DADOS USP	<a href="https://www.abcd.usp.br/noticias/dados-cientificos-como-construir-metadados-descricao-readme-dicionario-de-dados/">https://www.abcd.usp.br/noticias/dados-cientificos-como-construir-metadados-descricao-readme-dicionario-de-dados/</a>
<b>WEBINAR E EVENTOS EM FORMATO DE VÍDEO</b>	
RESEARCH DATA LIFECYCLE - UK DATA SERVICE	<a href="https://www.youtube.com/watch?v=-wjFMMQD3UA">https://www.youtube.com/watch?v=-wjFMMQD3UA</a>
I e II CORETRUSTSEAL CERTIFICATION (WEBINAR)	<a href="https://www.coretrustseal.org/why-certification/requirements/">https://www.coretrustseal.org/why-certification/requirements/</a>
FAIRSF AIR 2021 - FAIRSF AIR IN THE CONTEXT OF THE EOSC	<a href="https://www.youtube.com/watch?v=lpjw6vHCgSM&amp;list=PLXXwep9JCnSQzgfS4A-zkt7lEdhnapQ">https://www.youtube.com/watch?v=lpjw6vHCgSM&amp;list=PLXXwep9JCnSQzgfS4A-zkt7lEdhnapQ</a>
CORETRUSTSEAL CERTIFICATION	<a href="https://www.youtube.com/watch?v=gs9HTzxb0s">https://www.youtube.com/watch?v=gs9HTzxb0s</a>

Fonte: Elaborado pelo autor (2022).

### 3.2.3 Pesquisa experimental

No objetivo “Analisar o Dataverse enquanto plataforma para repositórios de dados de pesquisa;”, utilizar-se-á da pesquisa experimental, justamente por envolver uso de *software* e de tecnologia, assim como, o teste da plataforma Dataverse. Segundo Prodanov e Freitas (2013), na pesquisa experimental, o pesquisador procura refazer as condições de um fato a ser estudado, para observá-lo sob controle. Para tal, ele se utiliza de local apropriado, aparelhos e instrumentos de precisão, a fim de demonstrar o modo ou as causas pelas quais um fato é produzido, proporcionando, assim, o estudo de suas causas e seus efeitos. As pesquisas experimentais são mais frequentes nas ciências tecnológicas e nas ciências biológicas. Elas têm como objetivo demonstrar como e porque determinado fato é produzido. A escolha do *software*, neste trabalho, aconteceu pelos seguintes critérios:

- a) é um dos *softwares* mais encontrados nos trabalhos de revisão de literatura;
- b) a parceria com o IBICT para a promoção do *software* no Brasil foi o principal critério, devido à tradução da documentação e ao incentivo com editais de financiamento, este deverá ser o *software* mais utilizado no país;
- c) presente na maioria das instalações de repositórios de dados brasileiros;
- d) criado especificamente para o depósito e gestão de dados, assim, não necessita ser customizado para atender aos conjuntos de dados.

### 3.3 DESENHO DA PESQUISA

O desenho será realizado de acordo com cada objetivo específico.

Quadro 19 – Desenho da pesquisa

DESENHO DA PESQUISA		
OBJETIVO	TÉCNICA	FONTES
a) Levantar os fundamentos teóricos e recomendações relacionadas à gestão de dados e à implantação de repositórios de dados de pesquisa;	Técnica de levantamento bibliográfico e documental	Realizado no capítulo 2. Especificamente nas seções 2.2, 2.3, 2.5, 2.6, 2.7 e 2.8.
b) Modelar o serviço de suporte à gestão de dados de pesquisa para bibliotecas universitárias;	Pesquisa bibliográfica e documental	Realizado na seção 5.6 e corroborando com a implantação da gestão de dados na instituição 5.1, 5.2, 5.3 e 5.4.
c) Analisar o Dataverse enquanto plataforma para repositórios de dados de pesquisa;	Pesquisa experimental	Realizado no capítulo 4

d) Propor minuta de uma política de gestão de dados que contemple as boas práticas de gestão e de preservação de dados.	Pesquisa bibliográfica e documental	Atendido na seção 5.5.
--	-------------------------------------	------------------------

Fonte: Elaborado pelo autor (2022).



## 4 ANÁLISE DA PLATAFORMA DATAVERSE

Neste capítulo, apresenta-se a plataforma Dataverse, o histórico do *software*, os responsáveis pelo projeto, as funcionalidades e os teste na plataforma que ocorreram de março a julho de 2022. A proposta da pesquisa é experimentar e descrever a plataforma Dataverse do ponto de vista do bibliotecário, entender o ambiente e o depósito de dados no repositório, assim, cumpre-se o objetivo específico, “Analisar o Dataverse enquanto plataforma para repositórios de dados de pesquisa”. As habilidades e os conhecimentos sobre a gestão de dados de pesquisa foram descritas (seção 2.3), assim como, os conhecimentos sobre os repositórios e as exigências de sua institucionalização (da seção 2.4 a 2.7). Deve-se mencionar que a plataforma Dataverse permite implementar algumas exigências da literatura de gestão de dados, para que os dados cumpram os princípios FAIR, como uso de um identificador globalmente exclusivo e persistente, sendo também uma plataforma que facilita a obtenção de um repositório digital confiável (certificado), descrito na seção 2.8, por já possuir as funcionalidades exigidas no processo de certificação, o que outras plataformas que não foram desenvolvidas especificamente para os dados de pesquisa não conseguem implementar ou necessitariam de maiores customizações.

### 4.1 A PLATAFORMA DATAVERSE

A plataforma de *software* *The Dataverse Project* foi desenvolvida pelo Institute for Quantitative Social Science (IQSS), da Harvard University, em 2006, coordenado pelo pesquisador Gary King. Esse projeto nasceu a partir de uma proposta anterior entre o instituto e a biblioteca da universidade, o *Virtual Data Center* (VDC) de 1997.

Precusores do VDC datam de 1987, compreendendo entidades como *software* pré-web para transferir automaticamente informações de catalogação por [File Transfer Protocol] FTP para outros sites em todo o campus automaticamente em horários designados, e antes disso para um guia de *software* autônomo para dados locais (DATAVERSE, 2022).

Segundo Araújo, Arellano e Ferrer (2018, p. 6),

[...] a plataforma Dataverse é uma arquitetura de *software* livre para a publicação, citação, análise, preservação e reuso dos dados de projetos de pesquisa. Ela é dedicada ao compartilhamento, arquivamento e promove a referenciação de dados de pesquisa.

Um repositório do Dataverse é a instalação do software, que hospeda vários arquivos virtuais chamados coleções do Dataverse. Cada coleção do Dataverse contém conjuntos de dados e cada conjunto de dados contém metadados descritivos e arquivos de dados (incluindo documentação e código que acompanham os dados). Como método de organização, as coleções do Dataverse também podem conter outras coleções do Dataverse (DATAVERSE PROJECT, c2022; GABRIEL JÚNIOR *et al.*, 2021).

O IBICT realizou uma aliança com o projeto Dataverse, adotando o *software* para tradução e divulgação no Brasil<sup>24</sup>. De acordo com Araújo, Arellano e Ferrer (2018, p. 6), “[...] o Dataverse é composto por vários dataverses<sup>25</sup> e conjuntos de dados”. A partir da instalação principal do Dataverse, é criado um subdataverse dentro desta, que pode ser customizado para possuir novas exigências de metadados, designação de funções e acesso ao conjunto de dados, ou seja, os subdataverses podem funcionar de forma independente. Cada repositório Dataverse pode representar uma entidade organizacional e os subdataverses podem representar departamentos, laboratórios e grupos, sendo assim, uma organização hierárquica.

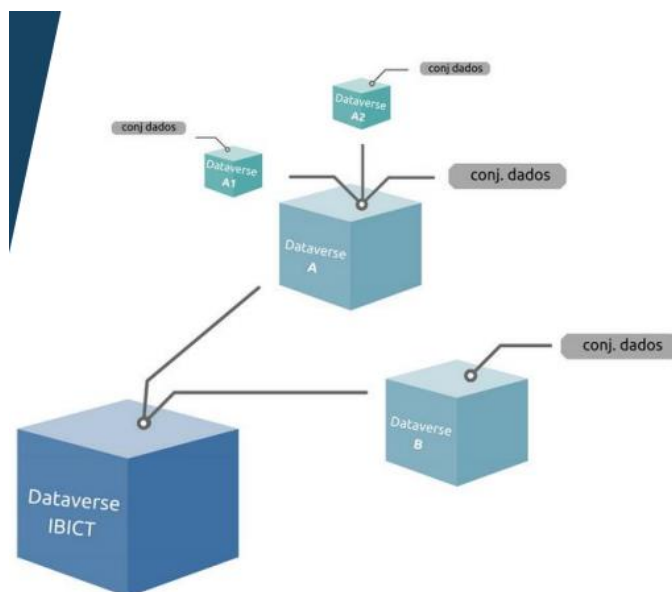
O Dataverse [...] possui recursos que possibilitam varia das configurações de ambientes de repositório de dados. A entidade dataverse é a estrutura que o software disponibiliza para representar organizações, grupos ou unidades. O software também é capaz de representar estruturas organizacionais hierárquicas, à medida que dataverses podem conter outros dataverses. Cada entidade dataverse contém datasets, que são as entidades que representam conjuntos de dados. O software dispõe de recursos para implementar entidades dataverses com políticas de funcionamento próprias e distintas. Uma entidade dataverse pode ser definida como se fosse um repositório independente, com todas as autorizações para gerenciamento e operação (como definir papéis, permissões e grupos; criar e gerenciar entidades dataverses, datasets arquivos etc.). Cada entidade dataverse pode ter marca personalizada (interface, logotipo) e interface configurada para descoberta de dados (ROCHA *et al.*, 2021, p. 8).

Os conjuntos de dados armazenados dentro de comunidades dataverses devem possuir: os metadados descritivos; os termos, como o glossário de termo; e o PGD, sendo as versões dos arquivos de dados (ROCHA *et al.*, 2021, p. 11). A cada conjunto de dados publicado é atribuído um *Digital Object Identifier* (DOI). Segundo Rocha *et al.* (2021, p. 17), a plataforma pode ser configurada via interface de programação (API) “para integração com o ambiente Archivematica, que assume a responsabilidade de preservar os *datasets* a longo prazo, enquanto o Dataverse assume a função de dar acesso aos dados”.

<sup>24</sup> A RDPBrasil e a RNP fizeram a tradução da versão atual do Dataverse.

<sup>25</sup> Em reunião entre CNPq, Ibict, RNP, FioCruz, UFRA, UFC, UFG, CBPF foi estabelecido a tradução para o Brasil “Comunidade Dataverse” quando se refere a dataverses criados dentro de instalações principais.

Figura 11 – Estrutura do *software* Dataverse



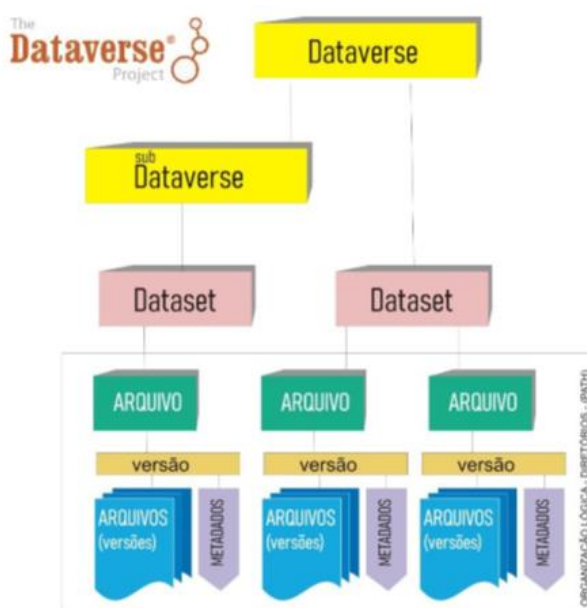
Fonte: Araújo, Arellano e Ferrer (2018, p. 7).

Após a parceria mencionada com o IBICT, este *software* possui documentação e guia de instalação disponível em português. E também vários relatórios<sup>26</sup> do projeto de pesquisa da Rede de Dados de Pesquisa Brasileira (RDP Brasil), compostos por pesquisadores da Universidade Federal do Rio Grande do Sul (UFRGS) em parceria com a Universidade Federal do Rio Grande (FURG), com coordenação executiva da Rede Nacional de Ensino e Pesquisa (RNP) e do IBICT. Os objetivos estratégicos do Dataverse Project ([2022], não paginado) são:

- Aumentar a adoção (usuários, repositórios do Dataverse, coleções do Dataverse, conjuntos de dados, periódicos);
- Desenvolver capacidade para lidar com dados sensíveis, em grande escala e streaming de nível 3;
- Expandir recursos de dados e metadados para disciplinas existentes e novas;
- Expandir características de arquivamento e preservação;
- aumentar as contribuições da comunidade de desenvolvimento de código aberto;
- melhorar UX e UI;
- Continuar a aumentar a qualidade do software.

<sup>26</sup> COLEÇÃO DE RELATÓRIOS: 2018-2020. Relatórios RDP - Dados de Pesquisa. Porto Alegre: CEDAP UFRGS, 2021. Disponível em: <https://cedap.ufrgs.br/jspui/handle/20.500.11959/1199>. Acesso em: 25 ago. 2021.

Figura 12 – Representação da estrutura hierárquica do Dataverse



Fonte: Rocha *et al.* (2021, p. 7).

#### 4.1.1 Perfis da plataforma DATAVERSE

O Dataverse permite definir níveis de permissões aos usuários, para isso, eles devem estar cadastrados no sistema, então o administrador/curador cadastra esse usuário e atribui a ele a permissão adequada. Os usuários do sistema são os utilizadores que vão fazer *download* e reusar os dados depositados, os pesquisadores que vão depositar seus dados para compartilhamento e os próprios administradores e curadores do repositório.

Para a definição de políticas, o software trabalha os conceitos permissão, papel e grupo. Permissão representa a autorização para realização de uma determinada funcionalidade do ambiente, como publicar um conjunto de dados. Papel corresponde a um conjunto de permissões, que caracteriza um perfil exercido por usuários na gestão ou operação do repositório. O ambiente já disponibiliza vários papéis (Administrador, Publicador de Dataverse, Curador de Dataverse, Editor de *Dataset*, etc.), mas novos papéis podem ser criados. Grupo representa um conjunto de usuários habilitados a atuar com determinados papéis (ROCHA *et al.*, 2021, p. 8).

Deve-se chamar atenção a três tipos de usuários, que deverão ser os mais comuns. Para Araújo, Arellano e Ferrer (2018, p. 14), existem os: (1) usuários com autorização que podem acessar e baixar um arquivo publicado; (2) os usuários sem autorização, que podem visualizar Dataverses da instalação e conjuntos de dados não publicados; e (3) os usuários externos, aqueles que não têm cadastro no sistema, podem visualizar dataverses e conjuntos de dados e baixar arquivos liberados.

Figura 13 – Permissões dos perfis Dataverse

	Administrador	Curador	Contribuidor	Usuário com autorização para acessar arquivos (Member)	Usuário sem autorização para acessar arquivos (File downloader)	Usuário externo
Adicionar dataverse	sim	sim				
Adicionar conjunto de dados	sim	sim				
Visualizar dados não publicados	sim	sim				
Visualizar conjunto de dados não publicado	sim	sim	sim	sim		
Baixar arquivo "restrito"	sim	sim	sim	sim	sim	
Editar dataverse	sim					
Editar conjunto de dados	sim	sim	sim			
Gerenciar permissões do dataverse	sim					
Gerenciar permissões do conjunto de dados	sim	sim				
Publicar dataverse	sim					
Publicar conjunto de dados	sim	sim				
Excluir dataverse	sim	sim	sim			
Visualizar dataverse e conjunto de dados publicados	sim	sim	sim	sim	sim	sim
Baixar arquivo "liberado"	sim	sim	sim	sim	sim	sim

Fonte: Araújo, Arellano e Ferrer (2018, p. 18).

#### 4.1.2 Metadados do Dataverse

Segundo Tamaro (2008, p. 212), “[...] o principal objetivo dos metadados numa biblioteca digital é auxiliar o usuário a encontrar a informação de que precisa e acessar o documento de modo mais simples e conveniente”. Para a autora, os metadados exercem uma função essencial no acesso à biblioteca digital para a integração, ou seja, identificação, localização e acesso.

O uso de diferentes tipos de padrões de metadados (padrões de estrutura, regras de conteúdo, marcação e empacotamento) garante:

- padrões de estrutura de metadados – representam uma estrutura consistente entre as entradas individuais, permitem a implementação de pesquisa de dados e o compartilhamento. Os padrões de estrutura hierárquica permitem que o contexto e o conteúdo sejam descritos;
- regras de conteúdo de metadados – permitem a entrada consistente de dados para a eficácia da pesquisa. Incluem os vocabulários e regras semânticas, arquivos de autoridade, dicionários de sinônimos, classificações e ontologias;
- padrões de marcação de metadados – garantem que os metadados sejam legíveis por máquina e que pesquisas automatizadas possam ser realizadas;
- padrões de empacotamento de metadados – definem os links entre objetos digitais e seus metadados enquanto vinculam os componentes em pacotes de arquivamento (ALMEIDA, 2019, p. 87-88).

Sobre os metadados da plataforma Dataverse:

[...] fornece metadados para conjuntos de dados nas ciências sociais usando o padrão Data Documentation Initiative (DDI). Com o tempo, o aplicativo Dataverse se expandiu para incluir também metadados e suporte a arquivos para: domínios adicionais, como astronomia e ciências biomédicas, além de aumentar a interoperabilidade com outros sistemas. [...] o processo de expansão do suporte a outros domínios para fins de interoperabilidade, descoberta, preservação e reutilização. Ele também fornecerá um gráfico visual descrevendo todas as APIs atualmente suportadas, padrões/esquemas de metadados (baseados em DDI Codebook, Dublin Core, DataCite 3.1, Observatório Virtual para dados de astronomia, e ISA-Tab para dados biomédicos), ontologias e tesouros; juntamente com o que está planejado para ser suportado no futuro (por exemplo, DCAT, RDF e schema.org) (CASTRO, 2016, não paginado).

### 4.1.3 Os identificadores persistentes

Os identificadores persistentes são códigos usados para reconhecer permanentemente objetos/arquivos em formato digital para que continuem ativos e acessíveis no ambiente. São identificadores para acessar objetos e metadados via serviços de identificadores globais, o Dataverse utiliza-se do *Digital Object Identifier (DOI)*, *Handle System*, HTTP e HTTPS, para acesso e transferência de informações entre o cliente e o repositório. E também possibilita que o identificador ORCID seja representado junto aos nomes de pessoas, como depositantes e autores de *datasets* (ROCHA *et al.*, 2021).

O DOI para dados de pesquisa é fornecido pela DataCite, uma organização global sem fins lucrativos. Segundo Wilkinson *et al.* (2016, p. 2), o DOI é:

[...] um código usado para identificar objetos (geralmente digitais) de forma permanente e estável. Os DOIs fornecem um mecanismo padrão para recuperação de metadados sobre o objeto e geralmente um meio de acessar o próprio objeto de dados.

O DOI é um identificador duradouro, compatível com o URN, mas não somente isso: é sobretudo um portal para a variedade de serviços associados ao uso de um objeto digital. Isso significa que a variedade de serviços possíveis para qualquer objeto digital, deve ser possível o acesso a várias fontes de informação sobre esse objeto, necessárias para os diversos serviços, isto é, para diversos sistemas de metadados (TAMMARO, 2008, p. 228).

O Handle é um identificador global exclusivo para os recursos eletrônicos e resolve o problema de URL. Segundo RNP (2022), apesar de imutáveis, os Handles permitem que as informações de localização e acesso aos objetos sejam alteradas. Cada identificador está associado a um único recurso e a organização que criou ou mantém o recurso. O Sistema Handle também sustenta a infraestrutura técnica do DOI, que é um tipo especial de Handle. O Handle System é administrado pela Corporation for National Research Initiatives (CNRI) que autorizou

a DONA Foundation, uma organização sem fins lucrativos, a assumir a responsabilidade pela evolução da arquitetura de objetos digitais do CNRI (HANDLE.NET, 2022, não paginado).

## 4.2 TESTE NA PLATAFORMA DATAVERSE

O começo dos testes ocorreu em 15 de março de 2022 na instalação (<https://dadospreservados.rnp.br/>), a versão v. 5.4.1, *Created using Sphinx 1.5.6*, com último *updated* em 13 abril de 2021. Na semana de 25 de abril de 2022, mudou-se para a instalação Demo do Dataverse (<https://demo.dataverse.org/dataverse/demo/>).

Devido à utilização de uma instalação já pronta, não será possível relatar sobre o processo e as dificuldades na instalação. Estes processos estão documentados em: Gabriel Júnior *et al.* (2019, 2020a, 2020b, 2020c, 2020d,) e Azambuja *et al.* (sem data), outras informações também estão acessíveis na página Suporte e Experimentação a e-Ciência, do Grupo de Trabalho em Rede de Dados de Pesquisa do Brasil (GT\_RDP Brasil). Assim, os testes basearam-se em buscar entender os seguintes itens na plataforma:

- criação de *login* e senha;
- criação do subdataverses “Pesquisa Mestrado PPGInfo UDESC Dataverse”;
- realização de testes de inclusão, edição e exclusão de dataverses, *datasets* e arquivos, o que permitiu a análise do *software* de forma sistêmica;
- teste de inclusão de usuários e suas permissões.

### a) Criar uma conta de usuário

De acordo com o Guia do Usuário Dataverse (2022), um usuário registrado no sistema, pode: (a) criar sua própria coleção de dados (se permitido) e a personalizar; (b) adicionar conjuntos de dados às coleções do Dataverse (se permitido); (c) contribuir para conjuntos de dados existentes (se permitido); (d) solicitar acesso a arquivos restritos (se permitido).

Os tipos de *logins* permitidos são: (a) nome de usuário/e-mail e senha, quando o usuário faz o cadastro do seu e-mail pessoal; (b) *login* institucional, quando o usuário entra via autenticação federada *shibboleth*; (c) via ORCID, quando o usuário faz o *login* a partir da sua conta; (d) Microsoft Azure AD; (e) GitHubGenericName; (f) Google. As opções de (b) a (f) são tipos de autenticações remotas. Existe a opção de converter *login* para outro tipo, por exemplo, um usuário cadastrado com e-mail e senha ser convertido para *login* institucional.

Figura 14 – Criando uma conta Dataverse

**Account Information**

**Username \*** Create a valid username of 2 to 60 characters in length containing letters (a-Z), numbers (0-9), dashes (-), underscores (\_), and periods (.)

**Password \*** Your password must contain:

- At least 6 characters (passwords of at least 20 characters are exempt from all other requirements)
- At least 1 character from each of the following types: letter, numeral

**Retype Password \***

**Given Name \***

**Family Name \***

**Email \***

**Affiliation \***

**Position \***

**General Terms of Use \*** There are no Terms of Use for this Dataverse installation.

I have read and accept the Dataverse General Terms of Use as outlined above.

**Create Account** Cancel

Fonte: Captura de tela da instalação Dados Preservados RNP (15 mar. 2022).

Para a criação de um novo usuário por e-mail e senha, solicita-se o preenchimento do nome do usuário, o sobrenome, o e-mail, a afiliação (instituição), a posição (cargo ocupado) e o aceite dos termos de uso (ver Figura 14). Após realizar esse cadastro, para realizar *login* só é necessário colocar o usuário e senha cadastrados (ver Figura 15).

Figura 15 – Conectar ao Dataverse

**Conecte-se**

Nome de usuário/e-mail

Senha

Conecte-se

Esqueceu sua senha?

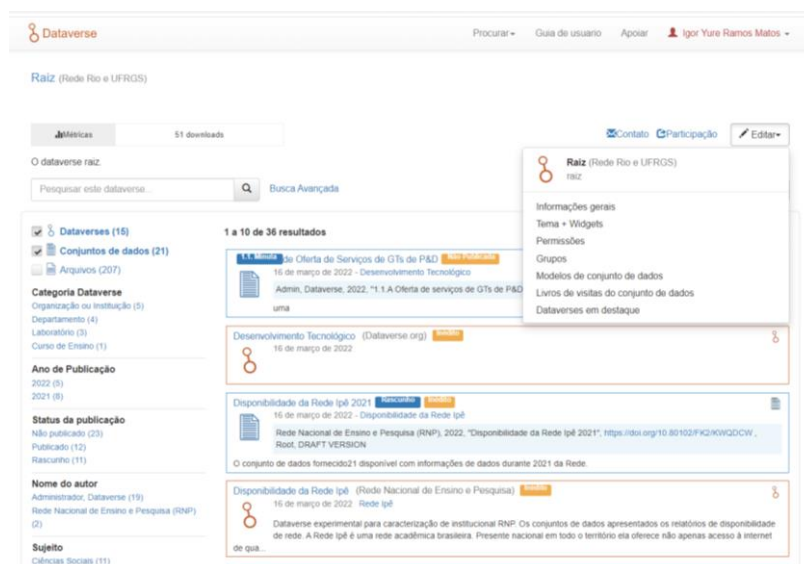
Inscreva-se para uma conta Dataverse.

Fonte: Captura de tela da instalação Dados Preservados RNP (15 mar. 2022).



Para os administradores, é permitido ter todas as funções do sistema, assim, consegue-se adicionar novos dataverses, subdataverses, conjuntos de dados, arquivo de dados, remover e alterar *link* de um *dataset*. A Figura 16 mostra as possibilidades que a plataforma oferece, como as informações gerais, a customização “tema + widgets”, a configuração de permissões etc.

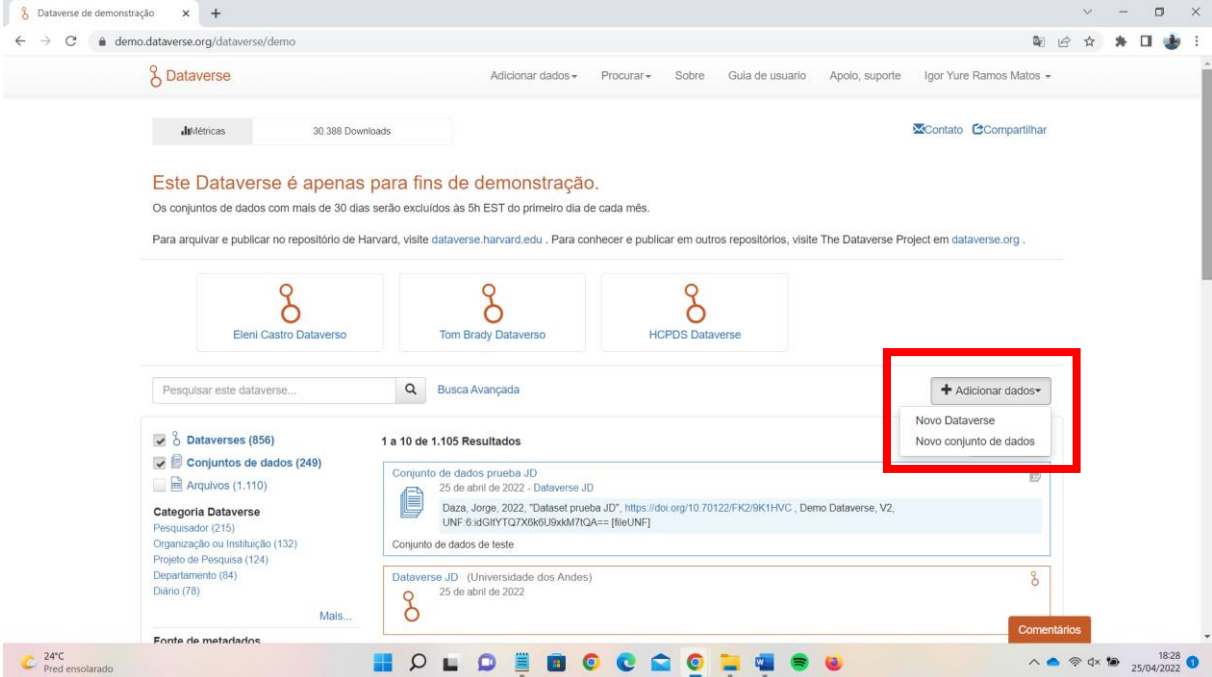
Figura 16 – Opções para configuração



Fonte: Captura de tela da instalação Dados Preservados RNP (18 mar. 2022).

Após acessar ao sistema, no botão “+ adicionar dados”, é possível inserir um novo subdataverse ou um conjunto de dados.

Figura 17 – Como adicionar DATAVERSE e conjuntos de dados



The screenshot shows the Dataverse demo website interface. At the top, there is a navigation bar with the Dataverse logo and links for 'Adicionar dados', 'Procurar', 'Sobre', 'Guia de usuario', 'Apoio, suporte', and 'Igor Yure Ramos Matos'. Below the navigation bar, there is a section with the text 'Este Dataverse é apenas para fins de demonstração.' and 'Os conjuntos de dados com mais de 30 dias serão excluídos às 5h EST do primeiro dia de cada mês.' There are three buttons for existing datasets: 'Eleni Castro Dataverse', 'Tom Brady Dataverse', and 'HCPDS Dataverse'. A search bar is present with the text 'Pesquisar este dataverse...' and a 'Busca Avançada' button. On the right side, a red box highlights a dropdown menu with the following options: '+ Adicionar dados', 'Novo Dataverse', and 'Novo conjunto de dados'. The main content area shows search results for '1 a 10 de 1.105 Resultados', including a dataset titled 'Conjunto de dados prueba JD' and 'Dataverse JD (Universidad de los Andes)'. The Windows taskbar at the bottom shows the date as 25/04/2022 and the time as 18:28.

Fonte: Captura de tela da instalação Demo Dataverse (25 abr. 2022.)

Figura 18 – Criando um novo Dataverse

Novo Dataverse

Dataverse de demonstração >

\*Asteriscos indicam campos obrigatórios

**Host Dataverse** ⓘ  
Demo Dataverse

**Nome do Dataverse\*** ⓘ  
Pesquisa Mestrado PPGInfo UDESC Dataverse

**Identificador\*** ⓘ  
https://demo.dataverse.org/dataverse/ppginfo

**Categoria\*** ⓘ  
investigador

**E-mail\*** ⓘ  
igoryure.rm@gmail.com

**Afiliação** ⓘ  
Universidade do Estado de Santa Catarina

**Descrição** ⓘ  
Este campo suporta apenas alguns Tags HTML.

**Campos de metadados**  
Escolha os campos de metadados a serem usados em modelos de conjunto de dados e ao adicionar um conjunto de dados a esse dataverse.

Use campos de metadados do Demo Dataverse

- Metadados de Citação (Obrigatório) [+][Ver campos](#)
- Metadados Geoespaciais [+][Ver campos](#)
- Metadados de Ciências Sociais e Humanas [+][Ver campos](#)
- Metadados de Astronomia e Astrofísica [+][Ver campos](#)
- Metadados de Ciências da Vida [+][Ver campos](#)
- Metadados do Diário [+][Ver campos](#)
- Metadados MRA [+][Ver campos](#)
- Metadados da Graduate School of Design [+][Ver campos](#)
- Alliance for Research on Corporate Sustainability Metadados [+][Ver campos](#)
- Metadados da Iniciativa de Replicação de Ciência Política [+][Ver campos](#)
- Metadados PSI [+][Ver campos](#)
- Metadados CHIA [+][Ver campos](#)
- Metadados Digaal [+][Ver campos](#)

**Navegar/Pesquisar Facetas**  
Escolha os campos de metadados a serem usados como facetas para navegar em conjuntos de dados e dataverses neste dataverse.

Use facetas de navegação/pesquisa do Demo Dataverse

Todos os campos de metadados

- Termo de Classificação do Tópico
- Linguagem
- Nome do produtor
- Data de produção
- Tipo de Contribuinte
- Nome do Contribuinte
- Agência de concessão de

**Selecionado**

- Nome do autor
- Afiliação do autor
- Sujeito
- Termo de palavra-chave
- Data de depósito

Copyright © 2022, o presidente e membros do Harvard College | Política de Privacidade

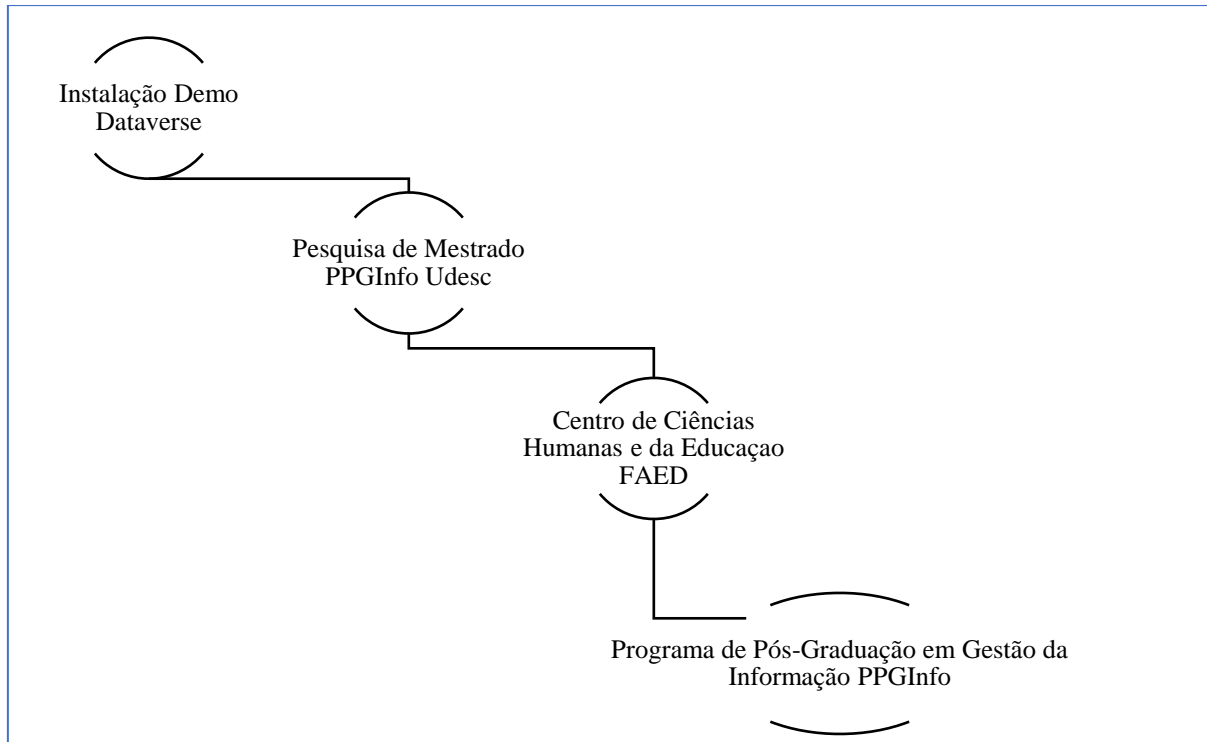
Desenvolvido por v. 5.10.1 build 807-b84672

The Dataverse Project

Fonte: Captura de tela da instalação Demo Dataverse (25 abr. 2022).

Ao criar um subdataverse dentro de uma instalação principal, o campo “host dataverse” (instalação principal), preenche: o nome desse novo dataverse, o identificador, qual será o *link* de acesso, a afiliação e a universidade ou o órgão que está fazendo essa instalação. Os subdataverses criados nos testes estão representados em ordem hierárquica na Figura 19.

Figura 19 – Hierarquia dos dataverses criados



Fonte: Elaborada pelo autor (2022).

Na Figura 20, mostra-se como adicionar conjuntos de dados e suas documentações.

Figura 20 – Dataverse novo

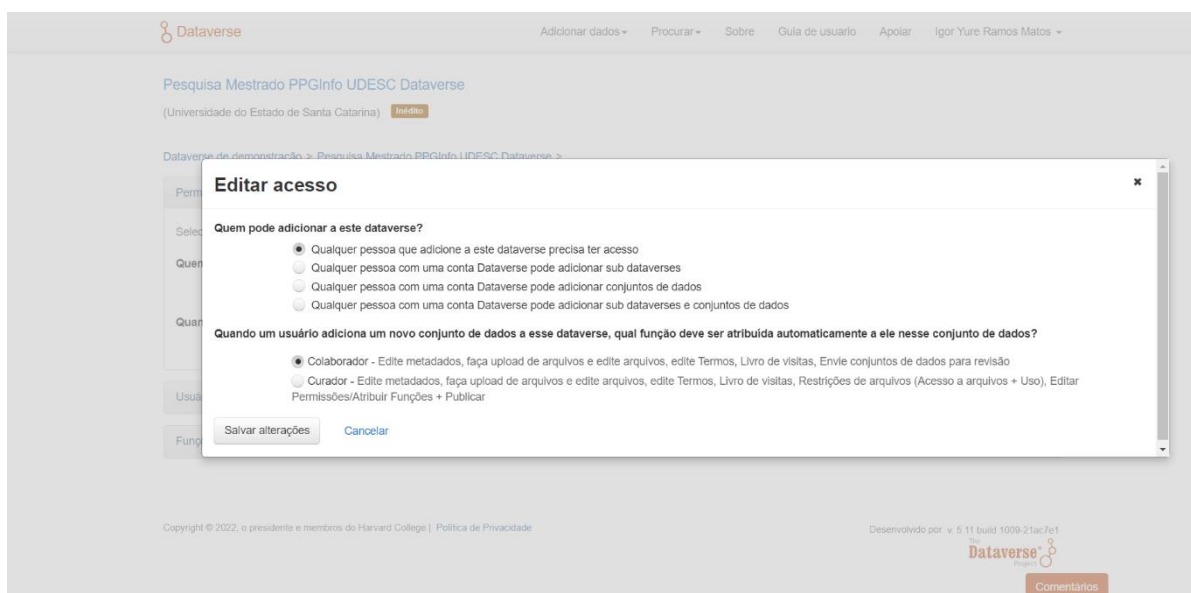
Fonte: Captura de tela da instalação Demo Dataverse (25 abr. 2022).

## b) Tipos de usuários

Sobre as permissões (ver Figura 13): o administrador é a pessoa que tem todas as permissões para dataverses, conjunto de dados e arquivos; o colaborador tem acesso a conjuntos de dados, é a pessoa que pode editar licenças e termos e enviá-los para revisão; o curador trabalha com conjuntos de dados, pode editar licenças, termos, permissões e publicar conjuntos de dados; o *dataverse and dataset creator*, criador de dataverse e *dataset* é a pessoa que pode adicionar subdataverses e conjuntos de dados dentro de um dataverse; o *dataverse creator*, ou seja, criador do conjunto de dados, pode adicionar conjuntos de dados em um dataverse; o *file downloader*, em português, baixador de arquivo, é o indivíduo que pode baixar um arquivo publicado; o membro é aquele que tem permissão para visualizar dataverses e conjuntos de dados não publicados.

Na aba “permissões”, no botão “editar acesso”, é possível permitir quem pode adicionar naquele dataverse/subdataverse. Para isso, basta o administrador definir previamente a função/permissão, ou seja, quem pode criar subdataverses, adicionar conjuntos de dados, qual a função que cada um pode exercer naquele novo conjunto de dados (Figura 21).

Figura 21 – Como atribuir permissões



Fonte: Captura de tela da instalação Demo Dataverse (10 jun. 2022).

Na aba “usuários/grupos” é possível ao administrador atribuir as funções acima descritas. Clicando em “atribuir funções a usuários/grupos” o administrador realiza a designação e designa funções aos usuários anteriormente cadastrados no sistema.

Figura 22 – Atribuindo funções e permissões no Dataverse

Copyright © 2022, o presidente e membros do Harvard College | Política de Privacidade

Desenvolvido por v. 5.10.1 build 907-b844672

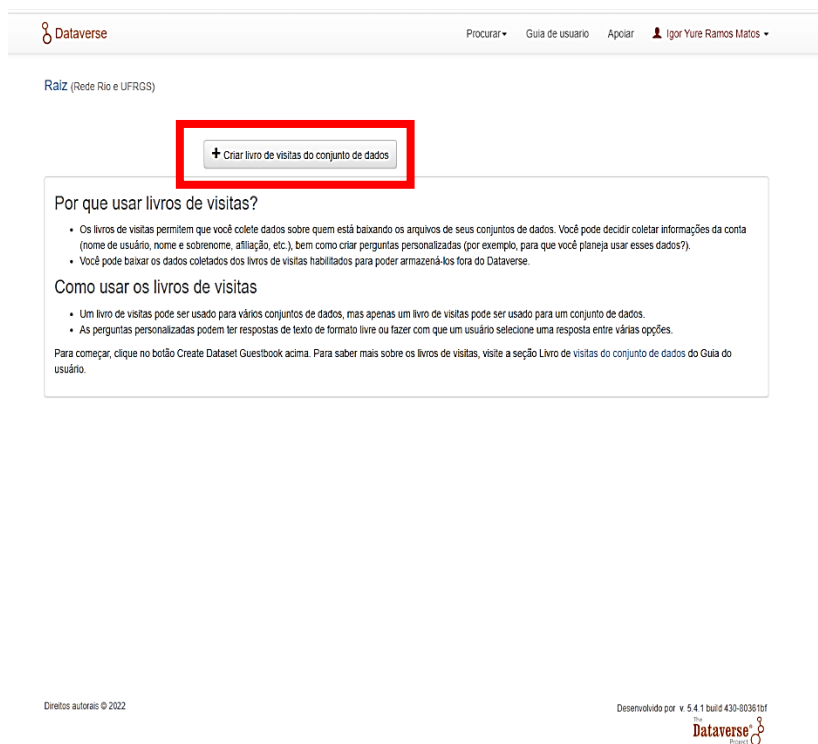
The Dataverse Project

Comentários

Fonte: Captura de tela da instalação Demo Dataverse (25 abr. 2022).

O livro de visitas permite o controle de quem consulta e faz *download* do conjunto de dados. É um recurso útil para se ter registro dos possíveis usos futuros dos dados. Para cada conjunto de dados é necessário um único livro de visitas, ou seja, cada “arquivo de dados/conjunto de dados” terá seu próprio registro de usuários que realizaram a visualização e o *download* daquele registro.

Figura 23 – Sobre o livro de visitas de um conjunto de dados



Fonte: Captura de tela da instalação Demo Dataverse (25 abr. 2022).

### c) Formato de citação

O projeto Dataverse padroniza a citação de seu conjunto de dados, isso, para que os pesquisadores obtenham créditos e reconhecimentos ao disponibilizarem seu trabalho, a citação é gerada automaticamente (DATAVERSE, 2022). Os campos da citação de dados, de acordo com Gabriel Júnior *et al.* (2021), são:

Quadro 20 – Modelo de citação Dataverse

<p><b>Autor(es)</b>, <b>Ano de publicação</b>, <b>Título do Dataset</b>, <b>Nome do repositório</b>,  <b>Versão</b>, <b>Identificador global persistente</b>.</p>
---

Fonte: Gabriel Júnior *et al.* (2021, [p. 7]).

Quadro 21 – Campos de citação de datasets no Dataverse

Campo	Descrição
<b>Autor(es)</b>	Nome dos autores;
<b>Ano de publicação</b>	Data da publicação dos dados de pesquisa;
<b>Título do Dataset</b>	Nome do <i>Dataset</i> definido pelo(s) autor(es);
<b>Nome do repositório</b>	
<b>Versão</b>	Identificação da versão dos dados disponibilizados;
<b>Identificador global persistente</b>	Número do DOI ou <i>Handle</i> utilizado para identificar os dados de pesquisa.

Fonte: Gabriel Júnior *et al.* (2021, [p. 7-8]).

Segundo o Projeto Dataverse (2022), o padrão de citação aqui definido oferece reconhecimento adequado aos autores, bem como, identificação permanente através do uso de identificadores globais e persistentes, no lugar de URLs, que podem mudar com frequência.

d) Interface de Aplicativos de Programação (API)

API significa *Application Programming Interface*, ou seja, Interface de Aplicativos de Programação. As APIs do *software* Dataverse permitem que você aproveite as integrações com outros *softwares* que podem ter sido configurados pelos administradores de sua instalação do Dataverse (DATAVERSE, 2021a).

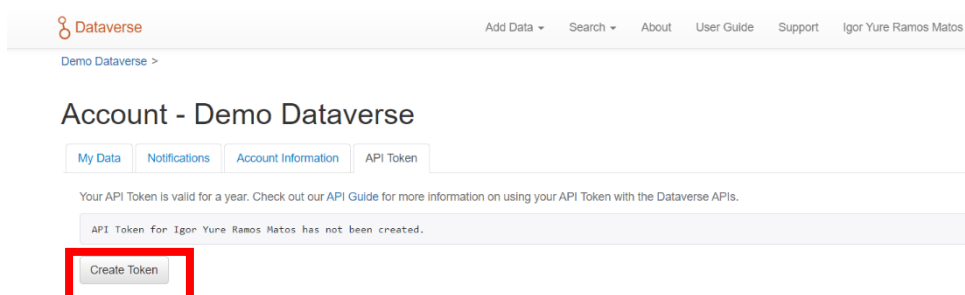
Alguns métodos e tecnologias garantem a interoperabilidade. Métodos como *Application Programming Interface* (API) proporcionam o intercâmbio de informações entre os repositórios de dados. As APIs são sistemas que permitem, por meio de um conjunto de rotinas e padrões de programação, que um aplicativo de *software* tenha acesso a outro *software* baseado na *web* e reutilize dados e/ou outras informações (SEMELER, 2017, p. 94).

O Dataverse permite que metadados e arquivos que compõem *datasets* sejam extraídos do ambiente via interface de programação (API). Isso possibilita que ações de preservação digital sejam realizadas por aplicativos externos ao Dataverse, permitindo a integração com o Archivematica, que assume a responsabilidade de preservar os *datasets* a longo prazo, enquanto o Dataverse responsabiliza-se pela função de dar acesso aos dados (ROCHA *et al.*, 2021).

Seguindo o Guia do Usuário (2022), o token de API é como uma senha, para algumas funcionalidades, por exemplo, ao depositar dados, é necessário um token de API para interagir com as APIs do *software* Dataverse. “Para criar seu token de API, clique no nome da sua conta na barra de navegação e selecione ‘Token de API’ no menu suspenso. Nesta aba, clique em ‘Criar Token’” (DATAVERSE, 2021b).



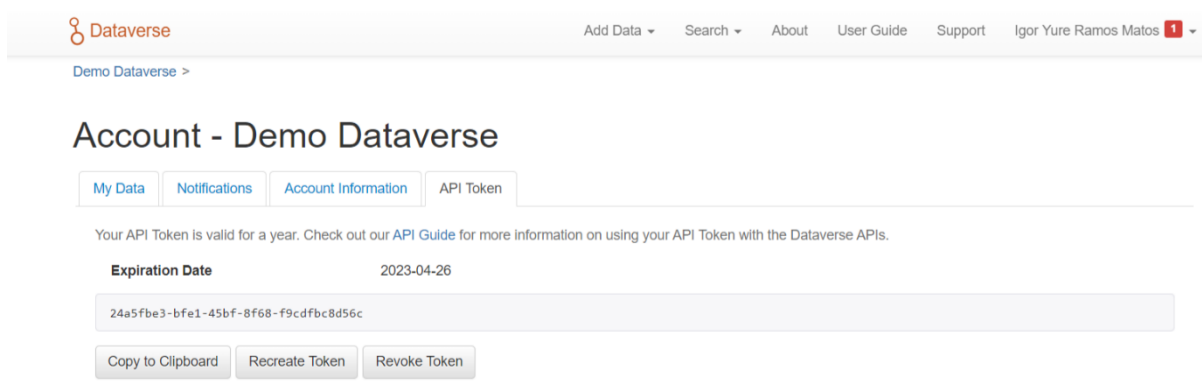
Figura 24 – Token de API



Fonte: Captura de tela da instalação Demo Dataverse (26 abr. 2022).

Segundo definição do Dataverse ([2021b]), a palavra *token* indica uma série de letras e números como c6527048-5bdc-48b0-a1d5-ed1b62c8113b. Qualquer pessoa que tenha seu *token* de API pode adicionar e excluir dados como você faria, portanto, você deve tratá-lo com o mesmo cuidado que uma senha. Conforme Rocha *et al.* (2021), o *software* disponibiliza uma API que permite o gerenciamento (submissão, criação, edição e remoção) de dataverses, *datasets* e arquivos, assim como, de usuários, papéis e grupos, possibilitando, ainda, que coleções sejam alimentadas a partir da colheita de metadados de dados armazenados em outros repositórios, via protocolos OAI-PMH.

Figura 25 – Token de API criado



Fonte: Captura de tela da instalação Demo Dataverse (26 abr. 2022).

As “ferramentas externas podem fornecer recursos adicionais que não fazem parte do próprio Software Dataverse, como visualizações de arquivos de dados, visualização e curadoria” (DATAVERSE, 2021b).

Quadro 22 – Ferramentas externas

Ferramentas externas			
Ferramenta	Modelo	Alcance	Descrição
Explorador de dados versão 2 do Dataverse Data Explorer	Explorar	Arquivo	O Data Explorer é um aplicativo da <i>web</i> que lista as variáveis em um arquivo de dados tabular e permite que os usuários visualizem categorias e estatísticas resumidas, vejam gráficos dos dados e realizem análises de tabulação cruzada. Este componente autônomo foi desenvolvido para complementar o Projeto Dataverse. O Data Explorer é integrado como parte das páginas <code>dataset.xhtml</code> e <code>file.xhtml</code> , revelando um botão “explorar” para arquivos de dados tabulares na instalação.
Whole Tale	Explorar	Conjunto de dados	Uma plataforma para a criação de pacotes de pesquisa reproduzíveis que permite aos usuários lançar ambientes de análise interativos em contêineres baseados em ferramentas populares como Jupyter e RStudio. Usando essa integração, os usuários do Dataverse podem iniciar os ambientes Jupyter e RStudio para analisar conjuntos de dados publicados.
Visualizadores de arquivos	Explorar	Arquivo	Uma coleção de visualizadores de arquivos de dados que estão em conformidade com a interface de ferramentas externas do Dataverse, originalmente desenvolvida pelo <i>Qualitative Data Repository</i> .
Ferramenta de curadoria de dados	Configurar	Arquivo	Para curadoria de dados adicionando rótulos, grupos, pesos e outros detalhes para ajudar na reutilização informada. Consulte o arquivo <code>README.md</code> em <a href="https://github.com/scholarsportal/Dataverse-Data-Curation-Tool">https://github.com/scholarsportal/Dataverse-Data-Curation-Tool</a> para obter as instruções de instalação.

Fonte: Adaptado de Dataverse (2021a).

Realizou-se a submissão de um conjunto de dados no formato de planilha e o glossário de termos (Figura 26). Percebe-se que ao preencher os metadados, a plataforma gerou um DOI de teste automaticamente. Segundo Gabriel Júnior (2020d, p. 5), “[...] para ativar o DOI no Dataverse, primeiro ative a versão de teste, depois de tudo testado, aplique as configurações de produção”. Portanto, numa plataforma Dataverse, pode-se gerar DOIs válidos automaticamente.

Figura 26 – Submissão de conjunto de dados

**Pesquisa Mestrado PPGInfo UDESC Dataverse**  
(Universidade do Estado de Santa Catarina) Unpublished

Dataverse de demonstração > Pesquisa Mestrado PPGInfo UDESC Dataverse >

**✓ Sucesso!** – Este conjunto de dados foi criado.

**! Info** – Esta versão preliminar precisa ser publicada. Quando estiver pronto para compartilhar, publique -o para que outras pessoas possam ver essas alterações.

## Replication Data for: Dados do artigo AtoZ

Rascunho não publicado

Ramos Matos, Igor Yure, 2022. "Replication Data for: Dados do artigo AtoZ", <https://doi.org/10.70122/FK2/PBZZER>, Demo Dataverse, DRAFT VERSION

Citar conjunto de dados [Saiba mais sobre os Padrões de Citação de Dados](#)

**Acessar conjunto de dados**  
Publicar conjunto de dados  
Editar conjunto de dados  
Contactar Proprietário  
Compartilhar

**Métricas do conjunto de dados**  
0 downloads

**Descrição**  
O conjunto de dados foi coletado das bases de dados: a) Base de Dados Referenciais de Artigos de Periódicos em Ciência da Informação (BRAPCI); b) Scientific Electronic Library Online (SciELO); c) Biblioteca Digital de Teses e Dissertações (BDTD); d) Library and Information Science Abstracts (LISA); e) Portal de periódicos CAPES. Pesquisou-se por: Repositório de dados e Ciência da Informação (2021-01-15)

**Sujeito**  
Ciências Sociais

**Palavra-chave**  
Repositório de dados, Gestão de dados de pesquisa, Ciência da Informação, Ciência Aberta

**Notas**  
Obteve-se na planilha final 127 artigos recuperados nas buscas nas bases da BRAPCI, Portal de periódicos CAPES, SciELO e LISA. Os dados foram coletados e salvos em formato de planilha ou documento. Em uma nova planilha, realizou-se a limpeza e organização dos metadados, manteve-se os seguintes campos, cada qual corresponde a uma coluna: título do artigo em português e inglês, autor(es), título do periódico, volume e número do periódico, International Standard Serial Number (ISSN), Digital Object Identifier (DOI) e Uniform Resource Locator (URL), ano, palavras-chave em português e em inglês e nome da base. Alguns dos metadados estavam incorretos, assim realizou-se a conferência manual, corrigindo, confirmando ou capturando-os novamente. Os artigos em inglês que não tinham palavras-chave em português, foram traduzidas utilizando a ferramenta tradutora do Google.

**Contrato de Licença/Usos de Dados**  
 CC0 1.0

arquivos **Metadados** **Termos** **Versões**

Pesquisar este conjunto de dados...

Fazer upload de arquivos

Filtrar por  
Tipo de arquivo: todos **Acesso: Todos**

Ordenar

1 a 2 de 2 arquivos **Editar arquivos** **Download**

	<b>Arquivo_teste_dataverse.xls</b> Planilha do MS Excel - 193,5 KB Depositado em 22 de julho de 2022 Conjunto de dados sobre a literatura dos repositórios de dados		
	<b>Revisão bibliográfica da Dissertação.xls</b> Planilha do MS Excel - 61,3 KB Depositado em 22 de julho de 2022 Conjunto de dados sobre a literatura dos repositórios de dados		

Copyright © 2022, o presidente e membros do Harvard College | Política de Privacidade

Desenvolvido por v. 5.11 build 1009-21ac7e1  
**The Dataverse Project**

**Comentários**

Fonte: Captura de tela da instalação Demo Dataverse (22 jul. 2022).

Sobre as licenças, “os Repositórios de Dados frequentemente utilizam opção *Creative Commons* CC BY1 como licença padrão” (GABRIEL JÚNIOR, 2020a, p. 4). Para alterar a licença de um *dataset*<sup>27</sup> é necessário: (a) acessar o conjunto de dados; (b) clicar em editar conjunto de dados, opção “termos”; (c) para mudar da licença CC0 1,0, escolher a opção “termos do conjunto de dados personalizados”; (d) serão habilitados os campos de metadados

<sup>27</sup> Desde a versão 5.10 é possível definir a CC-BY como default.

que permitem realizar restrições, permissões especiais etc. Para saber mais sobre a licença *Creative Commons* ver 2.3.5.1 e Open Data Commons 2.3.5.2.

Os testes realizados na plataforma ajudaram a entender como se dá o armazenamento dos conjuntos de dados, como funciona a criação de novas comunidades dataverses e como ocorre a atribuição de papéis e responsabilidades. Teve um caráter inicial e não abrangeu todas as funções e os testes completos da ferramenta, mas percebe-se que é uma plataforma fácil de ser entendida e utilizada por bibliotecários.

Assim, conhecer o que é exigido nos requisitos de certificação é importante, tanto para o trabalho de gestão de dados quanto para a parte tecnológica. A escolha da plataforma é imprescindível, pois, ela deve possuir os requisitos para implementar o armazenamento, a preservação, o versionamento, a possibilidade de formato no *download*, entre outras funcionalidades, como a plataforma Dataverse permite.

Interessante observar que a literatura revela que a menor dificuldade para se disponibilizar os dados de pesquisa é a questão referente à infraestrutura tecnológica, afinal, a capacidade de hardware está cada vez maior, assim como a velocidade de processamento; sendo essa é uma das características do big data. Em contrapartida, questões inerentes ao aspecto político e humano são as relacionadas pelos autores como as que geram dificuldades para se disponibilizar o dado (COSTA; CUNHA, 2019, p. 291).

Baseado nos testes da plataforma, na revisão de literatura e nos documentos da certificação CoreTrustSeal, percebeu-se que as maiores dificuldades consistem na institucionalização do serviço e no reconhecimento, por parte das instâncias superiores das universidades, da importância dos repositórios de dados. Fica mais evidente que até o momento (julho de 2022), no Brasil, a FAPESP e o Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) exigem a gestão de dados, o PGD dos pesquisadores e o depósito dos datasets em repositório de dados.

## 5 MODELO PARA A IMPLANTAÇÃO DE REPOSITÓRIO DE DADOS DE PESQUISA

Neste capítulo, baseado na revisão de literatura e nos testes realizado na plataforma Dataverse, apresenta-se uma proposta de implantação da gestão de dados nas universidades (Figura 27), os atores envolvidos (Figura 28), as responsabilidades dos atores (Quadro 23), dos pesquisadores (Quadro 24), dos bibliotecários (Quadro 25) e a estruturação de um serviço de suporte a dados de pesquisa (seção 5.6), com a finalidade de atingir o objetivo geral de “propor um modelo para implantação da gestão de dados de pesquisa nas universidades brasileiras”. De acordo com a fundamentação teórica e pelos modelos internacionais de gestão de dados, as bibliotecas universitárias são os órgãos que ficarão responsáveis pela curadoria e pela gestão dos repositórios de dados, afinal, são elas que realizam a gestão dos repositórios digitais que armazenam as publicações e a memória científica das universidades.

Elaborou-se uma minuta de uma política de gestão de dados (Seção 5.5) para apoio à implementação da gestão de dados nas universidades. Segundo Almeida (2005), “as políticas ou diretrizes são planos gerais de ação, guias genéricos que definem linhas mestras, orientam a tomada de decisão e dão estabilidade à organização”. A construção dos elementos da política foi baseada na revisão de literatura, principalmente, no *Requisitos de repositórios de dados confiáveis CoreTrustSeal: 2020–2022*, na tese de Almeida (2019), e na minuta da política de RI da UDESC ([2020]). A importância de se basear em critérios da certificação CoreTrustSeal se dá reconhecendo que o serviço busca cumprir com boas práticas de acesso aberto, de gestão e de preservação de dados. Deve-se mencionar que não existe um modelo formal ou diretrizes na literatura brasileira orientando como as instituições devem redigir esse documento.

A certificação CoreTrustSeal divide-se em três blocos de requisitos: infraestrutura organizacional (R1 a R6), gerenciamento de objetos digitais (R7 a R14) e tecnologia (R15 e R16), percebe-se que o menor bloco é o da tecnologia. Os critérios que demandam mais ações são: infraestrutura organizacional e gerenciamento de objetos digitais, ou seja, são requisitos e ações exigidas tanto para um serviço de suporte a dados na universidade, quanto na gestão e ciclo de vida dos dados pelos pesquisadores, na existência de políticas e o reconhecimento formal do repositório de dados na instituição.

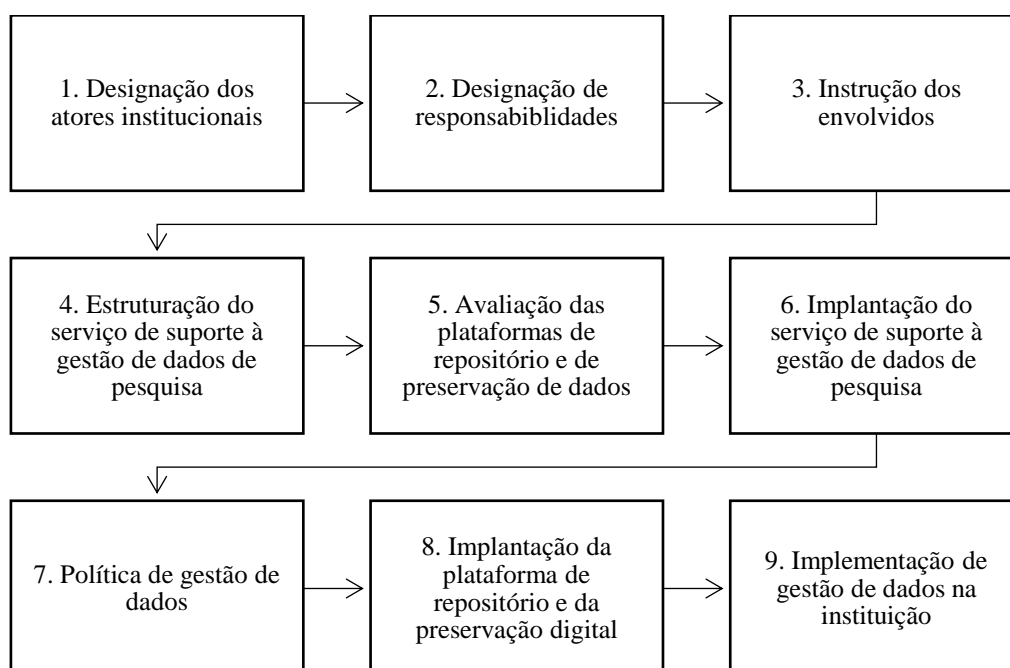
## 5.1 A IMPLANTAÇÃO DA GESTÃO DE DADOS NAS UNIVERSIDADES

A implantação da gestão de dados seguiu pautada nos modelos hierárquicos das universidades brasileiras, buscando o reconhecimento institucional, assim, propõe-se a execução com base nas seguintes fases:

- Primeira fase – designação dos autores envolvidos – define-se quais os órgãos, setores e pró-reitorias participarão do grupo de trabalho, são designados por portaria do reitor. Essa atribuição é definida pelos órgãos que já são responsáveis pelas pesquisas e/ou onde se produz dados de pesquisa; os responsáveis pela curadoria de repositórios digitais (logo também os repositórios de dados) e a secretaria responsável pela instalação e manutenção dos *softwares* e sistemas da universidade;
- Segunda fase – designação de responsabilidades – descreve-se, maneira geral, todas as responsabilidades dos atores definidos na primeira fase;
- Terceira fase – instrução dos envolvidos – essa refere-se à instrução dos membros do grupo de trabalho, para que todos os atores entendam a importância da gestão de dados. Só assim se garantirá a devida institucionalização e validação do assunto perante as instâncias superiores da universidade;
- Quarta fase – estruturação do serviço de suporte à gestão de dados de pesquisa – relaciona-se com a elaboração desse serviço dentro da biblioteca universitária (seção 5.6). A proposta aqui é separá-lo do serviço de referência/competência informacional, para a dedicação integral de tempo e de recursos, buscando profissionais capacitados e *experts* em gestão de dados de pesquisa e de repositórios de dados;
- Quinta fase – avaliação das plataformas de repositório e preservação de dados – testam-se as plataformas de repositórios mais conhecidas, suas funcionalidades, escolhendo-se assim a que melhor atende aos requisitos para a qualidade e a preservação dos conjuntos de dados. Deve-se escolher sempre uma plataforma de repositório que possibilite a integração com uma plataforma de preservação digital;
- Sexta fase – execução do suporte à gestão de dados de pesquisa – refere-se ao começo dos atendimentos e cursos no novo setor da biblioteca (quarta fase), constituindo-se pelo efetivo funcionamento do setor, com os cursos e atendimentos apresentados na seção 5.6;

- Sétima fase – política de gestão de dados – elaboração do documento que define as diretrizes da gestão de dados e do repositório de dados, podendo estar num único documento (conforme minuta elaborada na seção 5.5) ou em documentos separados;
- Oitava fase – implantação da plataforma do repositório e da preservação digital – após escolhidas as plataformas, realiza-se a instalação e a customização destas, sendo uma responsabilidade da Secretaria de Tecnologia da Informação, porém, realizar testes para verificar o funcionamento também será de responsabilidade do Serviço de Suporte à Gestão de Dados de Pesquisa;
- Nona fase – implementação da gestão de dados na instituição – relaciona-se com o cumprimento das fases anteriores, com documentos publicados (através de portarias) e institucionalizados, como a Política de Gestão de Dados e as Diretrizes para os Repositórios de Dados; a plataforma do repositório em funcionamento, armazenando os conjunto de dados; os pesquisadores, adotando o ciclo de vida, descrevendo os seus planos de gestão de dados, depositando e compartilhando dados FAIR; o serviço de suporte a dados de pesquisa funcionando, tanto o atendimento de orientação na biblioteca e nos laboratórios de pesquisa, quanto a parte de cursos de suporte e treinamento.

Figura 27 – Implantação da gestão de dados nas universidades

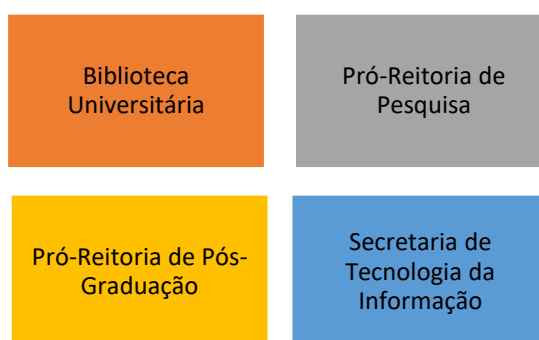


Fonte: Elaborada pelo autor (2022).

## 5.2 OS ATORES ENVOLVIDOS

Denominou-se de atores todos os órgãos envolvidos com a gestão de dados dentro das universidades, de acordo com o modelo educacional brasileiro de organização. Esses atores têm relação com as atividades e as responsabilidades que já lhe são atribuídas tanto na pesquisa quanto no atendimento à comunidade universitária. Eles deverão ser designados por portaria para compor o grupo de trabalho para a implementação da gestão de dados na instituição, poderão ser chamados de Comitê de Gestão de Dados, Grupo de Trabalho de Gestão de Dados, Comissão de Gestão de Dados, a denominação dependerá de como a instituição costuma chamar suas equipes de trabalho.

Figura 28 – Atores envolvidos na implementação da gestão de dados



Fonte: Elaborada pelo autor (2022).

Quadro 23 – Responsabilidade dos atores

<b>Biblioteca Universitária</b>	<b>Pró-Reitoria de Pesquisa</b>	<b>Pró-Reitoria de Pós-Graduação</b>	<b>Secretaria de Tecnologia da Informação</b>
<ul style="list-style-type: none"> <li>• Presidir o Comitê de Gestão de Dados;</li> <li>• Estruturar o setor/departamento de Serviço de Suporte a Dados de Pesquisa;</li> <li>• Implementar junto à comissão, as ações de gestão de dados;</li> <li>• Capacitar os bibliotecários que irão atuar neste novo setor de suporte a dados, desde cursos de extensão até licenças para mestrado e doutorado no tema;</li> <li>• O repositório garante que seus servidores/funcionários</li> </ul>	<ul style="list-style-type: none"> <li>• Integrar o Comitê de Gestão de Dados;</li> <li>• Realizar o levantamento dos laboratórios de pesquisa existentes na instituição;</li> <li>• Realizar o levantamento/censo dos tipos, formatos e volumes de dados produzidos nos laboratórios;</li> <li>• Incentivar que os professores realizem a gestão e o armazenamento de dados no repositório da instituição;</li> </ul>	<ul style="list-style-type: none"> <li>• Fazer parte do Comitê de Gestão de Dados;</li> <li>• Incentivar que os alunos de pós-graduação realizem a gestão de dados;</li> <li>• Estimular que os alunos façam o plano de gestão de dados de suas teses e dissertações;</li> <li>• Incentivar que os alunos da pós-graduação realizem os treinamentos oferecidos pelo Serviço de Suporte à Pesquisa, da Biblioteca.</li> </ul>	<ul style="list-style-type: none"> <li>• Fazer parte do Comitê de Gestão de Dados;</li> <li>• Hospedar a infraestrutura tecnológica do repositório de dados;</li> <li>• Instalar e customizar a plataforma de repositório;</li> <li>• Realizar as atualizações quando necessário;</li> <li>• Implementar <i>software</i> de preservação digital com modelo (OAIS);</li> <li>• Manter acesso contínuo ao repositório de dados da instituição;</li> </ul>



tenham acesso a treinamentos contínuos e desenvolvimento permanente.	<ul style="list-style-type: none"> <li>• Incentivar que os pesquisadores realizem os treinamentos e os cursos oferecidos pelo Serviço de Suporte à Pesquisa.</li> </ul>		<ul style="list-style-type: none"> <li>• Dar espaço em nuvem para os pesquisadores armazenem seus dados antes do depósito;</li> <li>• Fornecer serviços de armazenamento, segurança e <i>backup</i> de dados.</li> </ul>
--	---	--	--

Fonte: Elaborado pelo autor (2022).

### 5.3 GRUPO DE TRABALHO DE GESTÃO DE DADOS

Designado por portaria, este grupo de trabalho será composto dos atores envolvidos na implementação da gestão de dados: Biblioteca Universitária, Pró-Reitoria de Pesquisa, Pró-Reitoria de Pós-Graduação e Secretária de Tecnologia da Informação. Podem ser acrescentados ao grupo, professores que estudem a ciência aberta, a gestão de dados e os repositórios de dados, mesmo estes não fazendo parte dos setores mencionados acima. Esse grupo tem por objetivo a implementação da gestão de dados na universidade, sendo assim, tem função consultiva e deliberativa. Alguns dos resultados/documentos originados deste grupo deverão ser: (a) a Política de Gestão de dados da instituição; (b) as Diretrizes do Repositório de Dados e; (c) a implementação da plataforma de repositório de dados.

### 5.4 AS RESPONSABILIDADES DOS PESQUISADORES E BIBLIOTECÁRIOS

O pesquisador ou gestor do projeto será utilizado para denominar o indivíduo responsável pela coleta do conjunto de dados, podendo ser o coordenador do laboratório, o líder da pesquisa ou o professor assistente que acompanha ou realiza a coleta de dados. Os pesquisadores que realizam os estudos podem ser tanto indivíduos sozinhos (predominante nas ciências humanas), quanto dezenas de pessoas, vinculadas a determinado laboratório de pesquisa (ciências da saúde), ou mesmo, grupos independentes de pesquisa que utilizam do mesmo espaço e equipamentos.

Quadro 24 – Responsabilidades dos pesquisadores

<b>Coordenador do projeto</b>
<ul style="list-style-type: none"> <li>• Realizar o planejamento da pesquisa;</li> <li>• Organizar o ciclo de vida dos dados;</li> <li>• Elaborar e atualizar o Plano de Gestão de Dados;</li> <li>• Elaborar um “glossário de termos” para cada <i>dataset</i> ou arquivo de dados;</li> <li>• Coletar e/ou produzir dados: reunir, descrever, limpar, analisar e publicar os resultados;</li> <li>• Garantir a validade, a completeza, a integridade e a consistência dos dados;</li> <li>• Realizar a curadoria dos dados dentro dos laboratórios;</li> <li>• Realizar a preservação dos dados;</li> <li>• Salvar os arquivos de dados em formato aberto;</li> </ul>

- Utilizar padrão internacional de metadados;
- Identificar dados sensíveis e realizar suas adequações;
- Assegurar as questões éticas e de privacidade dos dados;
- Solicitar a retenção de dados;
- Depositar os dados, a documentação e descrever os metadados em repositórios de dados;
- Elaborar protocolos para seguir o ciclo de vida adotado;
- Criar medidas de segurança e de prevenção para evitar perdas de dados: acidentais, *hardware* ou *software*, humanos ou ambientais;
- Atribuir o tipo de licença aberta a coleção de dados;
- Armazenar e fazer *backup* dos dados e metadados;
- Citar os conjuntos de dados utilizados por outros pesquisadores.

Fonte: Elaborado pelo autor (2022).

O curador de dados é o profissional que deverá se especializar nas demandas de implementação da gestão de dados, ele tem a função de avaliar os pesquisadores/gestores do projeto e os alunos de pós-graduação, possibilitando que eles tratem adequadamente os seus dados.

Nesse sentido, conclui-se esta tese recomendando que os bibliotecários de dados deveriam trabalhar com repositórios de dados de pesquisa, auxiliando pesquisadores desde as primeiras fases da investigação científica, ajudando com o processo de documentação e garantindo que os dados de pesquisa sejam preservados, utilizáveis e reutilizáveis em longo prazo. Um bibliotecário de dados deve usar os valores fundamentais, os princípios éticos, as habilidades e os conhecimentos profissionais da Biblioteconomia para trabalhar com dados de pesquisa, principalmente em repositórios digitais (SEMELER, 2017, p. 143).

Essa função surgiu, segundo Semeler (2017, p. 17), como

[...] resultado do empenho de bibliotecários norte-americanos, britânicos e canadenses em criar serviços e produtos de divulgação, de consultoria, de gerenciamento, preservação e elaboração de esquemas de metadados, para efetiva incorporação de dados de pesquisa em coleções de documentos.

#### Quadro 25 – Responsabilidade dos bibliotecários

<b>Bibliotecário/Curador de dados</b>
<ul style="list-style-type: none"> <li>• Apoiar e auxiliar os pesquisadores: trabalhando junto com os estudantes;</li> <li>• Auxiliar na manipulação dos dados, participando da coleta, da organização, da classificação e da preservação dos dados;</li> <li>• Oferecer serviços de referência em gestão de dados;</li> <li>• Possuir capacidade de trabalhar com qualquer formato e tipo de dado (quantitativos e qualitativos);</li> <li>• Auxiliar na elaboração do plano de gestão de dados;</li> <li>• Experiência no preparo e na limpeza dos dados: ferramentas e <i>softwares</i>;</li> <li>• Conhecer a comunicação científica e seus fluxos;</li> <li>• Conhecer os processos dos comitês de ética;</li> <li>• Conhecer os métodos e as técnicas de pesquisa;</li> <li>• Oferecer serviços de referência e de assistência técnica em dados de pesquisa;</li> <li>• Ter conhecimento a respeito de propriedade intelectual: direitos autorais e licenças abertas;</li> <li>• Conhecer os tipos de licenças abertas, suas diferenças e aplicações;</li> <li>• Conhecer os padrões e os esquemas de metadados;</li> </ul>

- Ter conhecimento sobre identificadores persistentes;
- Conhecer os Princípios FAIR;
- Saber utilizar o diário de campo digital (*e-notebook*);
- Conhecer *softwares* de repositórios de dados: funcionalidades, diferenças, vantagens e desvantagens;
- Realizar a curadoria e o gerenciamento dos conjuntos de dados no repositório;
- Publicar os conjuntos de dados no repositório de dados;
- Estruturar o serviço de referência dentro da biblioteca: Serviço de Suporte à Gestão de Dados de Pesquisa;
- Elaborar junto ao Comitê de Gestão de Dados a Política de Gestão de Dados;
- Elaborar de manuais e tutoriais sobre padrões de metadados;
- Conhecer *softwares* de preservação digital;
- Dominar assuntos como as certificações existentes para repositórios: CoreTrustSeal, DIN 31644 (2012) e ISO 16363 (2012);
- Saber a diferença entre a Big Science e a Small Science;
- Possuir habilidade em gestão, supervisão e liderança de equipes;
- Oferecer serviço de referência e de citação de dados;
- Instruir os pesquisadores em escolher o modelo de ciclo de vida de dados mais adequado à área, ao tipo de dado etc.;
- Capacitar os pesquisadores em como descrever o PGD e as ferramentas que podem ser utilizadas;
- Conhecer os metadados administrativos, descritivos e estruturais;
- Realizar a migração dos dados para outros formatos (quando necessário);
- Realizar a migração por interoperabilidade entre repositórios (quando necessário).

Fonte: Adaptado de Almeida (2019) e Semeler (2017).

## 5.5 POLÍTICA DE GESTÃO DE DADOS

O esboço da minuta de uma política (Apêndice D) ocorreu na percepção de que a maior dificuldade encontrada não está na instalação e no uso das funcionalidades da plataforma, mas sim na fase de implantação da gestão de dados e na definição de uma política de gestão de dados na instituição/universidade. Não foi encontrado um modelo de política de gestão de dados que contemplasse todos os itens necessários, conforme a literatura de gestão de dados, como: ciclo de vida, curadoria, princípios FAIR, PGD, licenças etc. Portanto, utilizou-se dos trabalhos de: Angelaki e Papadopoulou (2018), Lazzari (2019), *OpenAire* (2022), *UK Research and Innovation* (2015), para realizar o esboço de uma **Política de Gestão de Dados de Pesquisa**, dividida em nove capítulos:

- CAPÍTULO I – DEFINIÇÃO
- CAPÍTULO II – DOS BENEFÍCIOS
- CAPÍTULO III – DAS RESPONSABILIDADES
- CAPÍTULO IV – DO DEPÓSITO DOS DADOS DE PESQUISA
- CAPÍTULO V – DA PROPRIEDADE INTELECTUAL
- CAPÍTULO VI – DO ARMAZENAMENTO DE DADOS
- CAPÍTULO VII – DOS METADADOS
- CAPÍTULO VIII – DO REUSO DOS DADOS
- CAPÍTULO IX – DAS PENALIDADES

De acordo com Costa (2017, p. 72),

[...] uma política baseada no princípio do compartilhamento de dados provavelmente se concentrará nas práticas-chave necessárias para fornecer acesso aos dados [...] uma política baseada na administração de dados se concentrará nas funções e responsabilidades envolvidas no gerenciamento de dados.

Neste modelo, inserimos em um único documento ações de gestão e de implementação das diretrizes do repositório, justamente para que a comunidade universitária e seus gestores consigam enxergar, no todo, as ações e as finalidades dessa gestão de compartilhamento e de reuso dos dados. Outro ponto foi vinculado à exigência de alguma plataforma para a preservação digital dos dados, sendo que ela deveria ser compatível com o *software* escolhido, portanto, esse fator deve ser levado em conta no momento de escolha do *software* do repositório de dados, ou seja, já realizando o planejamento e a concepção de ações de preservação digital.

## 5.6 SERVIÇO DE GESTÃO DE DADOS DE PESQUISA

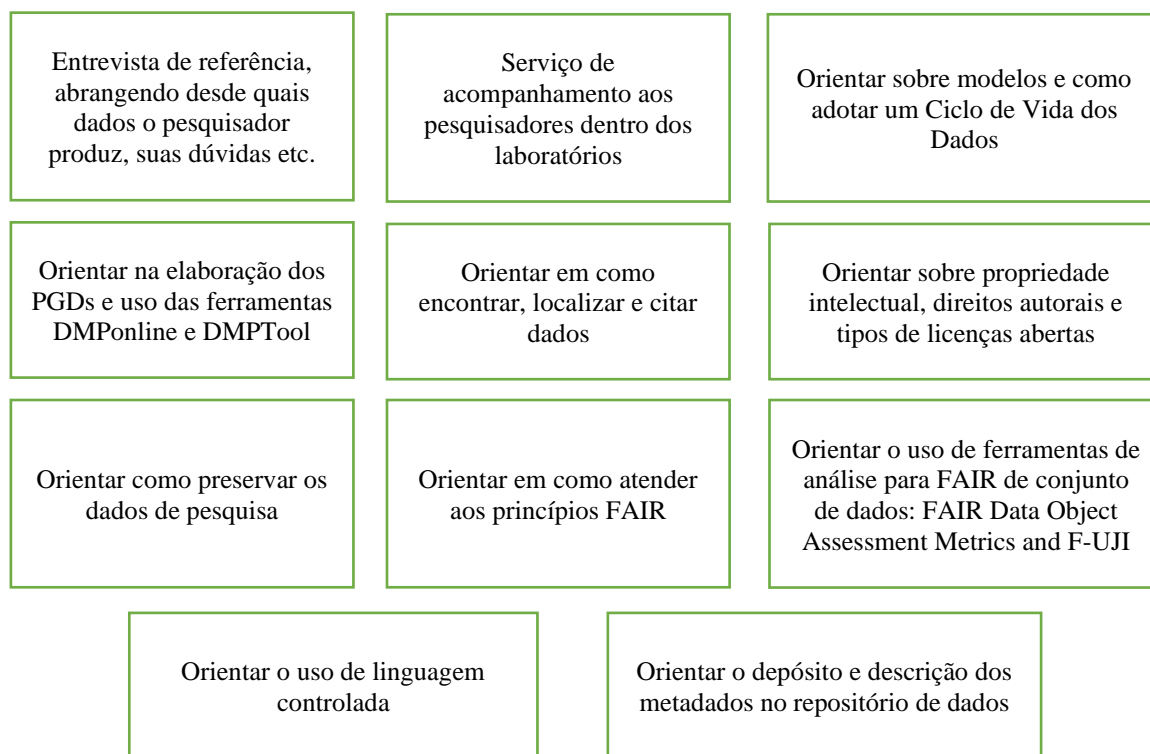
Para a estruturação de um Serviço de Suporte à Gestão de Dados de Pesquisa para bibliotecas universitárias brasileiras, conforme apresentado por Almeida (2019) em sua tese, deve-se atentar aos seguintes serviços: 1) serviços de gestão de dados: auxiliar os pesquisadores a realizarem o ciclo de vida, a descrição do PGD, o formato dos dados, a documentação, a descrição dos metadados e o armazenamento e; 2) serviços de curadoria: o armazenamento e a preservação dos dados a longo prazo em repositório de dados e a criação de produtos e/ou serviços de dados. Almeida (2019) divide os serviços em: infraestrutura, suporte e educacional. Porém, como propomos nesta pesquisa, a denominação utilizada no serviço de referência, dividiu-se, aqui, em: orientação, suporte e curadoria.

Segundo Figueiredo (1992), os subsídios para o projeto e a implantação de serviço de referência/informação devem ter por objetivo geral tornar possível e/ou facilitar o acesso às fontes informacionais relevantes aos programas de ensino, pesquisa e extensão. Por analogia, pode-se pensar em propor um objetivo para os serviços de suporte à gestão de dados, por exemplo: tornar possível o compartilhamento e o reuso de dados de pesquisa na instituição. Assim, apresentamos os serviços e os possíveis cursos, possibilitando que a instituição crie outros cursos, serviços e atendimentos de acordo com cada realidade institucional.

### 5.6.1 Orientação a dados de pesquisa

A orientação abrange os serviços relacionados à instrução no atendimento presencial aos usuários dentro do serviço de suporte a dados de pesquisa nas bibliotecas universitárias, podendo ser prestado de forma remota, se a biblioteca utilizar desse tipo de atendimento.

Figura 29 – Orientação para dados

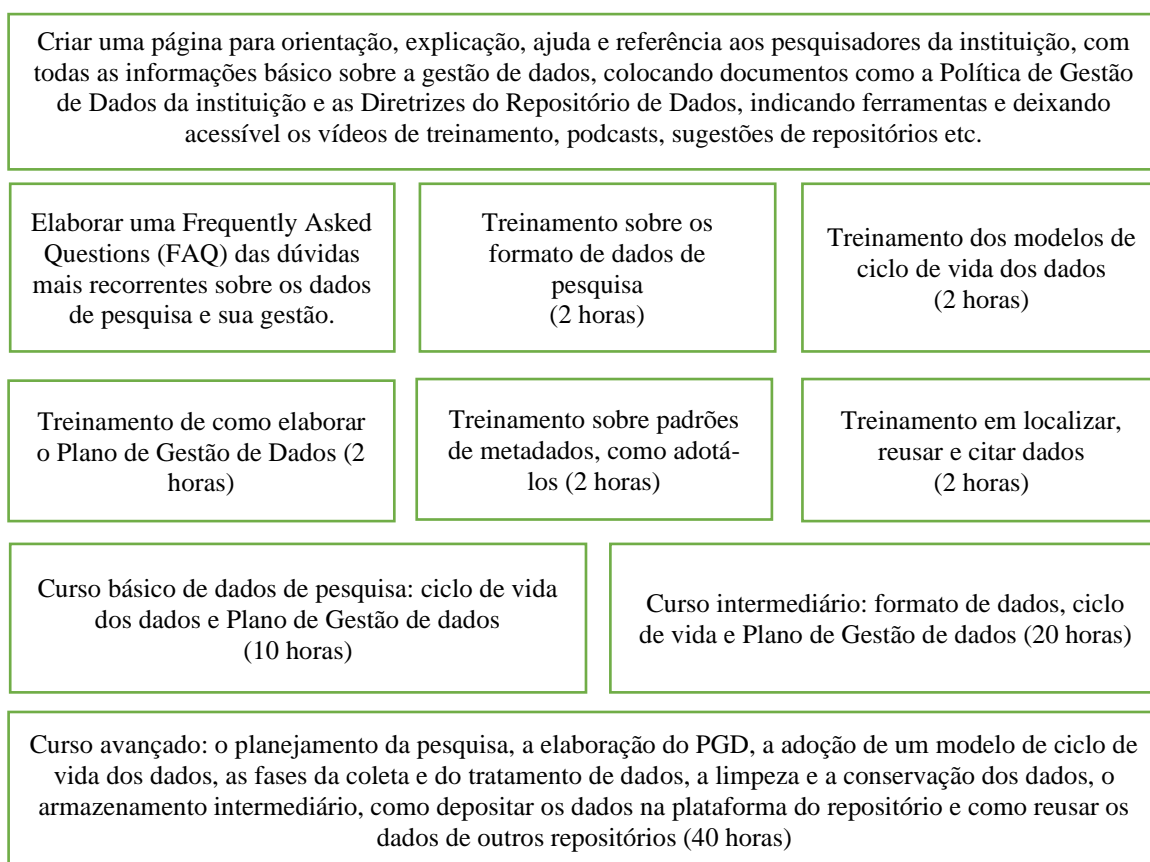


Fonte: Elaborada pelo autor, 2022.

### 5.6.2 Suporte a dados

O suporte abrange os serviços relacionados aos treinamentos, aos cursos e aos *workshops* prestados pelo setor. Será expandido, pois o suporte a dados exigirá que o bibliotecário trabalhe dentro dos laboratórios com os pesquisadores para os capacitar a: utilizar o ciclo de vida dos dados, descrever o PGD e os metadados de forma correta, preservar os conjuntos de dados, escolher melhor o formato para os dados e o uso de licenças abertas, bem como, o depósito em repositórios de dados. É realmente realizar o suporte à gestão de dados.

Figura 30 – Suporte a dados

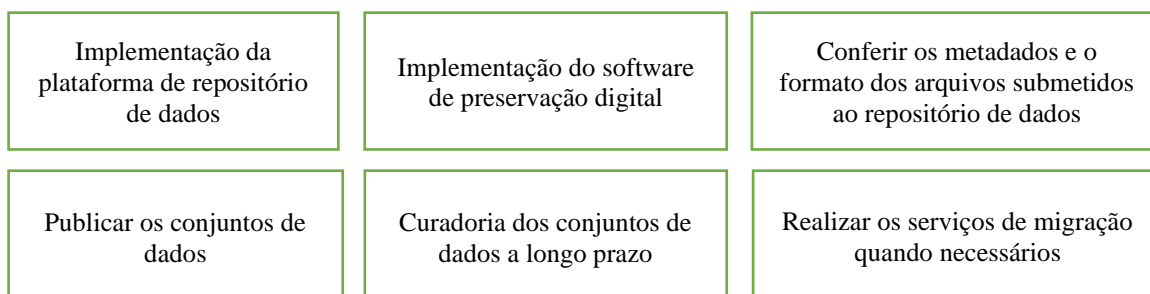


Fonte: Elaborada pelo autor, 2022.

### 5.6.3 Curadoria

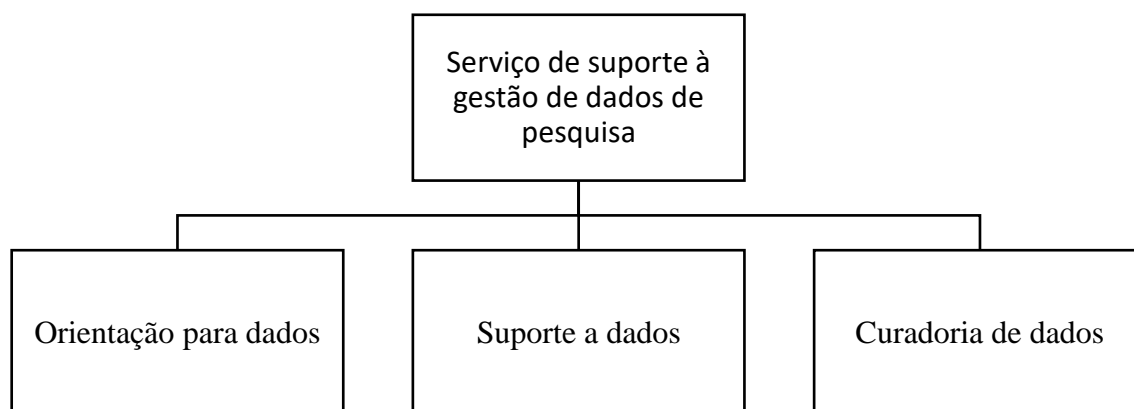
A curadoria abrange os serviços realizados na plataforma do repositório de dados, o tratamento dos metadados, a publicação dos conjuntos de dados, a migração de formatos (quando necessário) e o trabalho de preservação, ou seja, são as demandas executadas para que haja o acesso permanente aos dados de pesquisa.

Figura 31 – Curadoria de dados



Fonte: Elaborada pelo autor, 2022.

Figura 32 – Serviço de suporte a dados



Fonte: Elaborada pelo autor (2022).

## 6 CONSIDERAÇÕES FINAIS

Neste trabalho, percebeu-se que o quarto paradigma da pesquisa ou *e-Science*, caracterizado pelo uso intensivo de dados de pesquisa, transformou-os em insumos valiosos, novos *softwares* e novas possibilidades para a análise e o uso foram desenvolvidos, possibilitando novas descobertas científicas e inovações. Com o movimento da ciência aberta, as agências de fomento perceberam a necessidade de publicar os dados brutos também em acesso aberto, diminuindo os refinanciamentos e recoletas de informações parecidas, para isso, os pesquisadores precisam disponibilizá-los com qualidade, visando o reuso, atendendo, assim, aos princípios FAIR.

Esta pesquisa teve por objetivo propor um modelo para implantação da gestão de dados de pesquisa nas universidades brasileiras. Tal modelo pretende auxiliar na implementação da gestão de dados nas universidades brasileiras. O compartilhamento de dados, com a finalidade do reuso, é uma exigência das agências de fomento internacionais, que aos poucos, as agências brasileiras estão solicitando. Assim, as bibliotecas e os bibliotecários estão se capacitando para a gestão e a curadoria desses dados e seus repositórios. Entretanto, esses profissionais nunca trabalharam com conjuntos de dados, seus formatos, ciclos de vida e diferentes padrões de metadados, portanto, eles necessitam de capacitação. Para que isso seja possível, foram elencados e atendidos os seguintes objetivos específicos:

- a) Levantar os fundamentos teóricos e recomendações relacionadas à gestão de dados e à implantação de repositórios de dados de pesquisa; – foi realizada a apresentação dos dados de pesquisa com a exigência dos princípios FAIR para que sejam encontráveis, acessíveis, interoperáveis e reusáveis, relatou-se a necessidade de planejar a coleta dos dados adotando um modelo de ciclo de vida dos dados, buscando consistência e qualidade destes, há exigência que os pesquisadores descrevam os metadados e que façam um documento rico e detalhado sobre como a coleta aconteceu, ou seja, o PGD e a atribuição de licenças abertas para os *dataset*; em seguida, na parte de implantação e curadoria de repositórios, apresenta-se a exigência da certificação CoreTrustSeal para os repositórios de dados e aborda-se sobre a política de gestão de dados e o serviço de suporte à gestão de dados de pesquisa.
- b) Modelar o serviço de suporte à gestão de dados de pesquisa para bibliotecas universitárias – realizou-se a partir da literatura científica para estruturar o serviço e também, tendo como base a tese de Almeida (2019), porém, aqui inspirado no serviço



de referência já existente e solidificado nas bibliotecas brasileiras, buscando propor algo simples, dividindo em três categorias os serviços prestados: orientação, suporte e curadoria. Fez-se a sugestão de atendimentos, de tipos de cursos e de cargas horárias, assim como, a proposta de serviços realizados diretamente no tratamento dos dados e nas plataformas de repositório. O serviço é importante para a perenidade da gestão de dados.

- c) Analisar o Dataverse enquanto plataforma para repositórios de dados de pesquisa – na parte das técnicas, apresentamos os modelos de ciclo de vida dos dados que poderão ser adotados e as ferramentas para a descrição dos PGD; na parte tecnológica, o teste na plataforma Dataverse, percebendo que essa atende aos requisitos necessários para a qualidade e a preservação dos conjuntos de dados, como a adoção de identificadores persistentes, o versionamento e a rica descrição de metadados.
- d) Propor minuta de uma política de gestão de dados que contemple as boas práticas de gestão e de preservação de dados – para auxiliar a implantação nas instituições, baseado na literatura, principalmente nos critérios do CoreTrustSeal, elaborou-se uma minuta contendo diretrizes para o repositório e para gestão de dados. Destacando a necessidade da integração do *software* de preservação com a plataforma de repositório, a adoção de ciclo de vida dos dados, os planos de gestão de dados, a anonimização de dados sensíveis, as atribuições de licenças, ou seja, deixando, assim, um documento em forma de artigos, incisos, parágrafos e alíneas, facilitando e padronizando a leitura do documento.

Ainda, realizaram-se recomendações para o aprofundamento de novas pesquisas sobre os autores, coautores e colaboradores na captura dos conjuntos de dados, atribuindo os devidos créditos a quem realizou o trabalho, ou seja, o estudo dos direitos de propriedade intelectual dos dados. Outra proposta de estudo é o processo de obtenção de Repositórios de Dados Confiáveis (certificados). Espera-se que os repositórios de dados brasileiros busquem a certificação CoreTrustSeal, pois, promove a própria instituição ao reconhecimento perante a comunidade científica internacional.

O uso das ferramentas apresentadas neste trabalho, desenvolvidas no projeto FAIRsFAIR: Fair Aware, Fair Data Object Assessment Metrics e Automated FAIR Data Assessment Tool, constitui outra proposta para estudo, pois, contribui na qualidade dos

conjuntos, visando o reuso e a interoperabilidade, tanto por humanos quanto por máquinas, ou seja, o ideal para o compartilhamento dos dados na ciência aberta. Espera-se também que o modelo (seção 5.1) seja implantado nas Instituições de Ensino Superior (IES) para que possa ser validado e adequado à realidade de cada local.

Portanto, o serviço de repositório de dados deverá possuir mecanismos para a qualidade dos conjuntos de dados, ou seja, a adoção de boas práticas na gestão dos dados de pesquisa, desde a fase do planejamento, incentivando os produtores desses dados a adotarem o ciclo de vida, descreverem com metadados detalhados, elaborarem PGDs etc. Assim, a qualidade dos *datasets*, buscando o reuso, depende do comprometimento tanto dos pesquisadores quanto dos curadores dos repositórios de dados, denominados também de “bibliotecários de dados”. Novos eventos e ações estão acontecendo preparando esses bibliotecários para essa nova demanda profissional, por exemplo, a parceria do IBICT com o Dataverse Project (2015), o edital da RNP e IBICT, em 2021, incentivando a incubação de repositórios de dados em três universidades brasileiras. Os eventos promovidos da Rede Cariniana e Fiocruz, como: I Seminário Nacional de Repositórios Digitais (2022) e outros Webinars e palestras sobre o assunto.

## REFERÊNCIAS

- ALMEIDA, Fernanda Gomes. **Suporte à gestão de dados de pesquisa: uma ampliação dos serviços oferecidos pelas bibliotecas**. 2019. Tese (Doutorado em Gestão e Organização do Conhecimento) – Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2019. Disponível em: <http://hdl.handle.net/1843/31596>. Acesso em: 29 maio 2022.
- ALMEIDA, Maria Christina Barbosa de. **Planejamento de bibliotecas e serviços de informação**. 2. ed. ver. amp. Brasília: Briquet de Lemos, 2005.
- AMORIM, Ricardo Carvalho; CASTRO, João Aguiar; SILVA, João Rocha da; RIBEIRO, Cristina. A comparison of research data management platforms: architecture, flexible metadata and interoperability. **Universal Access in the Information Society**, [S.l.], v. 16, n. 4, p. 851-862, 2017. Disponível em: <https://link.springer.com/article/10.1007/s10209-016-0475-y>. Acesso em: 25 ago. 2021.
- ANGELAKI, Marina; PAPADOPOULOU, Elli. **Toolkit for policy makers on Open Access and Open Science**. 8. nov. 2018. Zenodo. Disponível em: <https://doi.org/10.5281/zenodo.2573797>. Acesso em: 20 ago. 2022.
- APPEL, André Luiz. **A e-Science e as atuais práticas de pesquisa científica**. 2014. Dissertação (Mestrado em Ciência da Informação) – Escola de Comunicação da Universidade Federal do Rio de Janeiro; Instituto Brasileiro de Informação em Ciência e Tecnologia, Rio de Janeiro, 2014. Disponível em: <http://ridi.ibict.br/handle/123456789/872>. Acesso: 29 maio 2022.
- ARAÚJO, Luiza Martins de Santana; ARELLANO, Miguel Ángel Márdero; FERRER, Igor Dias. **Guia para os usuários do repositório Dataverse do Ibict**. Brasília: IBICT, 2018.
- AZAMBUJA, Luís Alberto Barbosa *et al.* Acesso aberto a dados de pesquisa no Brasil: Dataverse: protocolo de migração.
- BARBIERI, Carlos. **Governança de dados: práticas, conceitos e novos caminhos**. Rio de Janeiro: Alta Books, 2019. 278 p.
- BELL, Gordon. Prefácio. *In*: HEY, Tony; TANSLEY, Stewart; TOLLE, Kristin (org.). **O quarto paradigma: descobertas científicas na era da e-Science**. Tradução Leda Beck. São Paulo: Oficina de textos, 2011. p. 11-15.
- BERMAN, Francine. Got data?: a guide to data preservation in the information age. **Communications of the ACM**, [New York], v. 51, n. 12, p 50–56, Dec. 2008, Disponível em: <https://doi.org/10.1145/1409360.1409376>. Acesso: 29 maio 2022.
- BERTIN, Patrícia Rocha Bello *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**, [S. l.], v. 15, n. 2, 2019. DOI: 10.18617/liinc.v15i2.4798. Disponível em: <http://revista.ibict.br/liinc/article/view/4798>. Acesso em: 31 maio 2022.

BORKO, Harold. Information science: what is it? **American Documentation**, v. 19, n. 1, 1968. Disponível em: <https://doi.org/10.1002/asi.5090190103>. Acesso: 17 ago. 2022.

BUCKLAND, Michael. Information as thing. **Journal of the American Society for Information Science (JASIS)**, v. 45, n. 5, p. 351-360, 1991. [Tradução livre Luciane Artêncio - Pós-Graduação em Ciência da Informação e Documentação – ECA/USP – 1º sem./2004].

CARIBÉ, Rita de Cássia do Vale. Comunicação científica: reflexões sobre o conceito. **Informação & Sociedade: Estudos**, v. 25, n. 3, p. 89-104, 2015. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/93078>. Acesso em: 12 ago. 2021.

CASTRO, Eleni. Community Data Repositories Working with The Library & University: a Harvard Dataverse use case. University of Massachusetts and New England Area Librarian e-Science Symposium. Disponível em: <https://doi.org/10.13028/4ytx-3753>. Acesso em: 17 ago. 2022.

CERVO, Amado Luiz; BERVIAN, Pedro Alcino. **Metodologia Científica**. 5. ed. São Paulo: Prentice Hall, 2002. 242 p.

CKAN. **CKAN**: The world's leading open-source data management system. [S. l.], 2021. Disponível em: <https://ckan.org>. Acesso em: 25 ago. 2021.

CONSELHO NACIONAL DE DESENVOLVIMENTO CIENTÍFICO E TECNOLÓGICO. Diretório dos grupos de pesquisa no Brasil. **Árvore do conhecimento**. Brasília, [201?]. Disponível em: <http://lattes.cnpq.br/web/dgp/arvore-do-conhecimento>. Acesso: 29 maio 2022.

CORETRUSTSEAL STANDARDS AND CERTIFICATION BOARD. CoreTrustSeal Trustworthy Data Repositories Requirements 2020–2022. [S.l.]: Zenodo, 20 Nov. 2019. Disponível em: <https://doi.org/10.5281/zenodo.3638211>. Acesso em: 02 jun. 2022.

CORETRUSTSEAL. **About**. Disponível: <https://www.coretrustseal.org/about/>. Acesso: 29 maio 2022.

CREATIVE COMMONS. **Sobre as licenças**. Mountain View: CC, [2017?]. Disponível em: <https://creativecommons.org/licenses/?lang=pt>. Acesso em: 19 maio 2021.

CUNHA, Murilo Bastos da. Metodologias para estudo dos usuários de informação científica e tecnológica. **Revista de Biblioteconomia de Brasília**, n. 10, v. 2, jul./dez. 1982. Disponível em: <https://periodicos.unb.br/index.php/rbbsb/article/view/30334>. Acesso: 29 maio 2022.

DATA CITATION SYNTHESIS GROUP. **Joint Declaration of Data Citation Principles**. Martone M. (ed.). San Diego, CA: FORCE11, 2014. Disponível em: <https://www.force11.org/group/joint-declaration-data-citation-principles-final>. Acesso em: 20 abr. 2022.

DATA OBSERVATION NETWORK FOR EARTH. **Data Life Cycle**. Disponível em: <https://old.dataone.org/data-life-cycle>. Acesso: 29 maio 2022.

DATAVERSE. **Guia de API**. [S. l.]: Harvard University, 2021a. [Dataverse v. 5.11]. Disponível em: <https://guides.dataverse.org/en/latest/api/index.html>. Acesso: 19 jun. 2022.

DATAVERSE. **Guia do usuário**. [S. l.]: Harvard University, 2021b. [Dataverse v. 5.11]. Disponível em: <https://guides.dataverse.org/en/latest/user/index.html>. Acesso: 19 jun. 2022.

DATAVERSE. **The Dataverse Project**: open-source research data repository software. [S. l.]: Harvard University, 2021. Disponível em: <https://dataverse.org/>. Acesso em: 26 ago. 2021

DECLARAÇÃO GO FAIR BRASIL (BR). Disponível em <https://www.go-fair-brasil.org/documentos>. Acesso: 19 jun. 2022.

DIGITAL CURATION CENTRE (DCC). **Curation Lifecycle Model**. [2022]. Disponível em: <https://www.dcc.ac.uk/guidance/curation-lifecycle-model>. Acesso em: 29 maio 2022.

DIGITAL PRESERVATION HANDBOOK. 2nd. [S. l.]: Digital Preservation Coalition, 2015. Disponível em: <https://www.dpconline.org/handbook>. Acesso: 29 maio 2022.

DMPONLINE. **Sobre o DMPonline**. Disponível em: <https://dmponline.dcc.ac.uk/>. Acesso: 19 jun. 2022.

DMPTOOL. **About**. Disponível em: <https://dmptool.org/>. Acesso: 19 jun. 2022.

DOCUMENTAÇÃO DDI 3.3. [S. l.: s. n.], 2020. Disponível em: <https://ddi-lifecycle-documentation.readthedocs.io/en/latest/User%20Guide/Introduction.html>. Acesso: 18 ago. 2022.

DUDZIAK, Elisabeth. **Dados de Pesquisa agora devem ser armazenados e citados**. São Paulo, Universidade de São Paulo, 2016. Disponível em: <http://www.sibi.usp.br/?p=6189>. Acesso em: 20 abr. 2022.

EDMUNDS, Rorie. **Webinar 1 Core Trust Seal Certification**. [S. l.: s. n.], 2018. Disponível em: <https://www.coretrustseal.org/why-certification/requirements/>.

EOSCSECRETARIAT.EU. **Working together for the realisation of EOSC: the infraEOSC 05 Collaboration Agreement**. 11 mar. 2020. Disponível em: <https://www.eoscsecretariat.eu/news-opinion/working-together-eosc-collaboration-agreement>. Acesso em: 06 jun. 2022.

ESPÍNDOLA, Priscilla Lüdtke *et al.* Governança de dados aplicada à ciência da informação: análise de um sistema de dados científicos para a área da saúde. **Revista Digital de Biblioteconomia & Ciência da Informação**, Campinas, v. 16, n. 3, p. 274-298, 2018. DOI: 10.20396/rdbci.v16i3.8651080. Acesso em: 05 jul. 2021.

FAIRSFAIR. **Support programme for Data Repositories**. Disponível em: <https://www.fairsfair.eu/application-results-open-call-data-repositories>. Acesso: 06 jun. 2022.

FAIRSFAIR. **The Project**: Fostering Fair Data Practices in Europe. Disponível em: <https://www.fairsfair.eu/the-project>. Acesso: 19 jun. 2022.

FAPESP. **Open Science @FAPESP**. São Paulo: Fapesp, [2019?]. Disponível em: <https://www.fapesp.br/openscience/>. Acesso em: 01 nov. 2020.

FORCE11: the future of research communications and e-scholarship. About FORCE11. Disponível em: <https://force11.org/info/about-force11/>. Acesso em: 29 maio 2022.

FREIRE, Gustavo Henrique de Araújo. Ciência da informação: temática, histórias e fundamentos. **Perspectiva em Ciência da Informação**, Belo Horizonte, v. 11 n. 1, p. 6-19, jan./abr. 2006.

FREITAS, Marília; LEITE, Fernando Cesar de Lima. Proposição de diretrizes para o depósito da produção científica em repositórios institucionais de acesso aberto baseada na visão de diferentes atores do sistema de comunicação científica. **Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação**, Florianópolis, v. 23, n. 53, p. 96-109, 2018. Disponível em: <https://www.brapci.inf.br/index.php/res/v/39627>. Acesso em: 31 mar. 2020.

GABRIEL JUNIOR, Rene Faustino *et al.* **Acesso aberto a dados de pesquisa no Brasil: dataverse: documentação de instalação e configuração – Ubuntu**, 2021. Disponível em: <https://cedap.ufrgs.br/jspui/handle/20.500.11959/1202>. Acesso em: 19 jun. 2022.

GABRIEL JUNIOR, Rene Faustino *et al.* **Acesso aberto a dados de pesquisa no Brasil: dataverse: licenças de uso**. 2020a. Disponível em: <http://hdl.handle.net/20.500.11959/1260>.

GABRIEL JUNIOR, Rene Faustino *et al.* **Acesso aberto a dados de pesquisa no Brasil: dataverse: customizações e traduções, [RELATÓRIO]**, 2020b. Disponível em: <http://hdl.handle.net/20.500.11959/1273>.

GABRIEL JUNIOR, Rene Faustino *et al.* **Acesso aberto a dados de pesquisa no Brasil: ativação do Handle**. 2020c. Disponível em: <http://hdl.handle.net/20.500.11959/1266>

GABRIEL JUNIOR, Rene Faustino *et al.* **Acesso aberto a dados de pesquisa no Brasil: Dataverse Documentação: Ativação do DOI**. 2020d. Disponível em: <http://hdl.handle.net/20.500.11959/1259>.

GABRIEL JUNIOR, Rene Faustino *et al.* **Acesso aberto a dados de pesquisa no Brasil: dataverse: Google Analytics no Dataverse**. 2019. Disponível em: <http://hdl.handle.net/20.500.11959/1203>.

GANNON, Dennis; REED, Dan. O paralelismo e a nuvem. *In*: HEY, Tony; TANSLEY, Stewart; TOLLE, Kristin (org.) **O quarto paradigma: descobertas científicas na era da e-Science**. Tradução Leda Beck. São Paulo: Oficina de textos, 2011. Cap. 16, p. 147-150.

GO FAIR. **Fair principles**. [S. l.: s. n.], 2022. Disponível em: <https://www.go-fair.org/fair-principles/>. Acesso: 19 jun. 2022.

GOMES, Anilza Rita de Souza; ROSA, Flávia Goullart Mota Garcia. Análise das políticas de funcionamento de repositórios institucionais brasileiros. **Ponto de Acesso**, v. 11, n. 1, p. 81-94, 2017. Disponível em: <https://www.brapci.inf.br/index.php/res/v/67146>. Acesso em: 31 mar. 2020.

GRAY, Jim. [Transcrição de Palestra em Montain View, Califórnia em 11 de janeiro de 2007]. TOLLE, Kristin; TANSLEY, Stewart; HEY, Tony. Jim Gray e a eScience: um método científico transformado. *In*: HEY, Tony; TANSLEY, Stewart; TOLLE, Kristin (org.) **O quarto paradigma: descobertas científicas na era da e-Science**. Tradução Leda Beck. São - Paulo: Oficina de textos, 2011. p. 17-29.

HENNING, Patrícia Corrêa *et al.* Desmistificando os princípios FAIR: conceitos, métricas, tecnologias e aplicações inseridas no ecossistema dos dados FAIR. **Pesq. Bras. em Ci. da Inf. e Bib.**, João Pessoa, v. 14, n. 3, p. 175-192, 2019. ISSN: 1981-0695. DOI: <https://doi.org/10.22478/ufpb.1981-0695.2019v14n3.46969>.

HUIGEN, Frans; VON STEIN, Ilona. **How is FAIRsFAIR supporting the repositories selected for CoreTrustSeal Certification?**. 2020. Disponível em: <https://fairsfair.eu/articles-publications/how-fairsfair-supporting-repositories-selected-coretrustseal-certification>. Acesso: 05 jun. 2022.

KUNH, Thomas S. **A estrutura das revoluções científicas**. Tradução Beatriz Vianna Boeira e Nelson Boeira. 13. ed. São Paulo: Perspectiva, 2018. (Série Debates, v. 115). Edição comemorativa dos 50 anos da publicação com ensaio introdutório de Ian Hacking. Título original: *The Structure of Scientific Revolutions*. ISBN 978-85-273-0111-4.

KURAMOTO, Hélio. A informação científica e o seu Acesso Livre: que direção o Brasil está adotando?. **Bibliotecas Universitárias: pesquisas, experiências e perspectivas**, [Belo Horizonte], v. 2, n. esp., 2015. Disponível em: <https://www.brapci.inf.br/index.php/res/v/17062>. Acesso em: 27 mar. 2020.

L'HOURES, H.; KLEEMOLA, M.; DE LEEUW, L. **CoreTrustSeal: from academic collaboration to sustainable services**. IASSIST Quarterly, [S. l.], v. 43, n. 1, p. 1–17, 2019. DOI: 10.29173/iq936. Disponível em: <https://iassistquarterly.com/index.php/iassist/article/view/936>. Acesso em: 14 mar. 2022.

LAZZARI, Letícia. **Políticas para gestão da produção científica da Universidade do Estado de Santa Catarina**. 2019. 127 p. Dissertação (Mestrado em Gestão da Informação) - Centro de Ciências Humanas e da Educação, Universidade do Estado de Santa Catarina, Florianópolis, 2019.

LEITE, Fernando César Lima *et al.* **Boas práticas para a construção de repositórios institucionais da produção científica**. Brasília: Ibict, 2012. Disponível em: <https://livroaberto.ibict.br/handle/1/703>. Acesso em: 10 jan. 2021.

LETROUIT, Carole *et al.* **A Place des bibliothèques universitaires dans le développement de la science ouverte**. [Paris]: Inspection générale de l'Éducation, du Sport et de la Recherche (IGÉSR), 2021. 66 p. Disponível em: <https://www.ouvrirelascience.fr/la-place-des-bibliothèques-universitaires-dans-le-developpement-de-la-science-ouverte/>. Acesso em: 05/07/2022.

MACHADO, Jorge. Dados abertos e ciência aberta. *In*: ALBAGLI, Sarita; MACIEL, Maria Lucia; ABDO, Alexandre Hannud (org.). **Ciência aberta, questões abertas**. Brasília: IBICT; Rio de Janeiro: UNIRIO, 2015. Disponível em: <http://livroaberto.ibict.br/handle/1/1060>. Acesso em: 21 nov. 2020. ISBN 978-85-7013-109-6. p. -227

- MARCONI, Marina de Andrade; LAKATOS, Eva Maria. **Técnicas de pesquisa:** planejamento e execução de pesquisas, amostragens e técnicas de pesquisa, elaboração, análise e interpretação de dados. 7. ed. São Paulo: Atlas, 2008. 277 p. ISBN 9788522451524.
- MARRA, Patrícia dos Santos Caldas. O papel das bibliotecas universitárias na comunicação científica: um estudo sobre os repositórios institucionais. **Encontros Bibli:** Revista Eletrônica de Biblioteconomia e Ciência da Informação, Florianópolis v. 17, n. esp., p. 174-194, 2012. Disponível em: <https://www.brapci.inf.br/index.php/res/v/39435>. Acesso em: 31 mar. 2020.
- MEDEIROS, Claudia Bauzer. Gestão de dados científicos: da coleta à preservação. **SciELO em Perspectiva**, jun. 2018. Disponível em: <https://blog.scielo.org/blog/2018/06/22/gestao-de-dados-cientificos-da-coleta-a-preservacao/>. Acesso: 17 maio 2022.
- MONTEIRO, Elizabete Cristina de Souza de Aguiar. **Direitos autorais nos repositórios de dados científicos:** análise sobre os planos de gerenciamento dos dados. 2017. 115 f. Dissertação (Mestrado em Ciência da Informação) – Faculdade de filosofia e Ciências, Universidade Estadual Paulista, Marília, 2017. Disponível em: <http://hdl.handle.net/11449/149748>. Acesso em: 20 out. 2020.
- MONTEIRO, Elizabete Cristina de Souza de Aguiar; SANT’ANA, Ricardo César Gonçalves. Repositórios de Dados Científicos na Infraestrutura de Pesquisa: adoção dos princípios FAIR. **Ciência da Informação**, Brasília, v. 48, n. 3 (Supl.), p. 347-353, set./dez. 2019. Disponível em: <https://revista.ibict.br/ciinf/article/view/4878>. Acesso em: 19 ago. 2022.
- MORENO, Fernanda Passini. Repositórios de dados de pesquisa na Espanha: breve análise. **Encontros Bibli:** revista eletrônica de Biblioteconomia e Ciência da Informação, Florianópolis, v. 23 n. 53, 2018. DOI: <https://doi.org/10.5007/1518-2924.2018v23n53p52>.
- MUELLER, Suzana Pinheiro Machado. Literatura científica, comunicação científica e ciência da informação. In: TOUTAIN, Lídia Maria Batista Brandão (org.). **Para entender a ciência da informação**. Salvador: EDUFBA, 2007. ISBN 978-85-232-0477-8. p. 125-144.
- OLIVEIRA, Adriana Carla Silva de; SILVA, Edilene Maria da. Ciência aberta: dimensões para um novo fazer científico. **Informação & Informação**, Londrina, v. 21, n. 2, p. 5-39, dez. 2016. DOI <http://dx.doi.org/10.5433/1981-8920.2016v21n2p5>. Disponível em: <http://www.uel.br/revistas/uel/index.php/informacao/article/view/27666>. Acesso em: 28 out. 2020
- OPENAIRE. Toolkit for policy makers on Open Science and Open Access: Open Science Policy Checklist for Research Performing Organisations (RPOs). Disponível em: <https://www.openaire.eu/open-science-policy-checklist-for-research-performing-organisations>. Acesso em: 20 ago. 2022.
- OPEN KNOWLEDGE FOUNDATION. Open Data Commons. **Legal tools for open data**. [Londres, 2022]. Disponível em: <https://opendatacommons.org/faq/licenses/>. Acesso de: 29 maio 2022.
- OPEN KNOWLEDGE FOUNDATION. Open Data Commons. **Open Definition 2.1**. [Londres, 2022]. Disponível em: <https://opendefinition.org/od/2.1/en/>. Acesso: 29 maio 2022.



PAGANINE, Lucas Nóbrega.; AMARO, Bianca. Características dos Repositórios de Dados Científicos no Brasil. **BIBLOS**, [s. l.], v. 34, n. 1, p. 176–188, 2020. DOI: 10.14295/biblos.v34i1.11132. Disponível em: <https://periodicos.furg.br/biblos/article/view/11132>. Acesso em: 12 ago. 2021.

PAVÃO, Caterina Groposo *et al.* **Acesso aberto a dados de pesquisa no Brasil**: políticas para repositórios de dados de pesquisa. 2019. Disponível em: <http://hdl.handle.net/20.500.11959/1263>. Acesso: 10 ago. 2021.

PAVÃO, Caterina Groposo; ROCHA, Rafael Port da; GABRIEL JUNIOR, Rene Faustino. Proposta de criação de uma rede de dados abertos da pesquisa brasileira. **RDBCI: Revista Digital de Biblioteconomia e Ciência da Informação**, Campinas, SP, v. 16, n. 2, p. 329–343, 2018. DOI: 10.20396/rdbci.v16i2.8651180. Disponível em: <https://periodicos.sbu.unicamp.br/ojs/index.php/rdbci/article/view/8651180>. Acesso em: 20 jun. 2022.

PRODANOV, Cleber Cristiano; FREITAS, Ernani Cesar de. **Metodologia do trabalho científico**: métodos e técnicas da pesquisa e do trabalho acadêmico. 2. ed. Novo Hamburgo: Feevale, 2013.

RE3DATA. **Registry of Research Data Repositories**. [S. l.], 2021. Disponível em: <https://doi.org/10.17616/R3D>. Acesso em: 15 fev. 2021.

UK RESEARCH AND INNOVATION. Guidance on best practice in the management of research data. [Jul.2015]. Disponível em: <https://www.ukri.org/publications/guidance-on-best-practice-in-the-management-of-research-data/>. Acesso em: 20 ago. 2022.

RIBEIRO, Cristina *et al.* **Os Repositórios de Dados Científicos**: estado da Arte. Projecto RCAAP D24 – Relatório, 2010. Disponível em: <https://repositorio-aberto.up.pt/handle/10216/23806>. Acesso: 10 ago. 2021.

ROBREDO, Jaime. **Da Ciência da Informação revisitada aos sistemas humanos de informação**. Brasília, DF: Thesaurus; SSRR Informações, 2003.

ROCHA, Rafael Port da *et al.* Análise dos sistemas DSpace e Dataverse para repositórios de dados de pesquisa com acesso aberto. **Revista Brasileira de Biblioteconomia e Documentação**, São Paulo, v. 17, p. 1-25, 2021. ISSN: 1980-6949. Disponível em: <https://rbbd.febab.org.br/rbbd/article/view/1572>. Acesso em: 19 jun. 2022.

SALARELLI, Alberto. A informação Digital: primeira parte. *In*: TAMMARO, Anna Maria; SALARELLI, Alberto. **A Biblioteca Digital**. Brasília: Briquet Lemos, 2008. p. 3-107.

SANCHEZ, Fernanda Alves; VECHIATO, Fernando Luiz; VIDOTTI, Silvana Aparecida Borsetti Gregório. Encontrabilidade da informação em repositórios de dados: uma análise do DataONE. **Informação & Informação**, Londrina, v. 24, n. 1, p. 51-79, jan./abr. 2019. ISSN 1981-8920. DOI: <http://dx.doi.org/10.5433/1981-8920.2019v24n1p51>. Disponível em: <http://www.uel.br/revistas/uel/index.php/informacao/article/view/30725>. Acesso em: 26 nov. 2020.

SARACEVIC, Tefko. A natureza interdisciplinar da ciência da informação. **Ciência Da Informação**, Brasília, v. 24, n. 1, 1995. Disponível em: <https://doi.org/10.18225/ci.inf.v24i1.608>. Acesso: 19 jun. 2022.

SARACEVIC, Tefko. Ciência da Informação: origem, evolução e relações. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 1, n. 1, p. 4-62, jan./jun. 1996. Disponível em: <https://periodicos.ufmg.br/index.php/pci/article/view/22308>. Acesso: 19 ago. 2022.

SAYÃO, Luís Fernando; SALES, Luana Farias. A ciência invisível: revelando os dados da cauda longa da pesquisa. *In*: ENCONTRO NACIONAL DE PESQUISA E PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO, 19., 2018, Londrina. p. [Anais...]. [S.l: s. n], 2018, p. 4180-4199. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/103678>. Acesso em: 20 jun. 2022.

SAYÃO, Luís Fernando; SALES, Luana Farias. A grande e a pequena Ciência: análise das diferenças na gestão de dados de pesquisa. **Informação & Sociedade: estudos**, [S. l.], v. 29, n. 3, 2019. Disponível em: <https://periodicos.ufpb.br/ojs2/index.php/ies/article/view/47615>. Acesso em: 18 ago. 2022.

SAYÃO, Luís Fernando; SALES, Luana Farias. Algumas considerações sobre os repositórios digitais de dados de pesquisa. **Informação & Informação**, v. 21, n. 2, p. 90-115, 2016. DOI: 10.5433/1981-8920.2016v21n2p90 Acesso em: 20 jun. 2022.

SAYÃO, Luís Fernando; SALES, Luana Farias. **Guia de gestão de dados científicos para bibliotecários de pesquisadores**. Rio de Janeiro: CNEN, 2015. Disponível em: <http://www.aben.com.br/noticias/guia-de-gestao-de-dados-de-pesquisa-para-bibliotecarios-e-pesquisadores>. Acesso em: 21 nov. 2020.

SEABRA, Carlos. Conteúdos digitais na internet e os direitos do público. *In*: DOWBOR, L.; SILVA, H. (org.). **Propriedade intelectual e direito à informação**. São Paulo: EDUC, 2014. Disponível em: [http://dowbor.org/blog/wp-content/uploads/2013/03/14Livro\\_Propriedade-intelectual.pdf](http://dowbor.org/blog/wp-content/uploads/2013/03/14Livro_Propriedade-intelectual.pdf). Acesso em: 22 nov. 2020.

SEMELER, Alexandre Ribas. **Ciência da informação em contextos de e-science: bibliotecários de dados em tempos de Data Science**. Tese (doutorado em Ciência da Informação) – Universidade Federal de Santa Catarina, Centro de Ciências da Educação, Programa de Pós-Graduação em Ciência da Informação, Florianópolis, 2017.

SILVA, Fabiano Couto Corrêa da. **Gestão de dados científicos**. Rio de Janeiro: Interciência, 2019. 128 p.

SILVA, Fabiano Couto Corrêa da. **Plano de Gestão de Dados**. Curso de Extensão Plano de Gestão de Dados: conceitos, usos e vantagens. [Curitiba]: UFPR virtual; PPGGI/UFPR, 2021. Aula 2. Slides.

STRASSER, Carly *et al.* **Primer on Data Management: what you always wanted to know**. California: CDL, 2012. Disponível em: <http://escholarship.org/uc/item/7tf5q7n3#page-1>. Acesso em: 20 maio 2021.

TAMMARO, Anna Maria. A biblioteca Digital: segunda parte. *In*: TAMMARO, Anna Maria; SALARELLI, Alberto. **A Biblioteca Digital**. Brasília: Briquet Lemos, 2008. p.111-339.

TARGINO, Maria das Graças. Comunicação Científica: uma revisão de seus elementos básicos. **Informação & Sociedade**, João Pessoa, v. 10, n. 2, 2000. Disponível em: <https://periodicos.ufpb.br/ojs/index.php/ies/article/view/326>. Acesso: 02 out. 2020.

UK DATA SERVICE. **Research Data lifecycle**. Disponível em: <https://ukdataservice.ac.uk/learning-hub/research-data-management/>. Acesso: 19 jun. 2022.

UNIVERSIDADE DE SÃO PAULO. Pró-Reitoria de Pesquisa. **Gestão de dados científicos**. [2020]. Disponível em: <http://prp.usp.br/gestao-de-dados-cientificos/?codmnu=9979>. Acesso em: 15 dez. 2020.

UDESC. Centro de Ciências Humanas e da Educação. **Programa de Pós-Graduação em Gestão da Informação**. Disponível em: <https://www.udesc.br/faed/ppginfo/areadeconcentracao>. Acesso em: 20 ago. 2021.

UDESC. Minuta de RESOLUÇÃO CONSUNI XX/2020. Institui o Repositório Institucional da Universidade do Estado de Santa Catarina (UDESC), o Comitê Gestor e a Política de Gestão. [documento em elaboração].

VERBURG, Maaïke *et al.* **D4.3 Report on the certification support and guidance for repositories and reviewers (V1.0\_DRAFT)**. [S. l.: s. n.], 2021. Disponível em: <https://doi.org/10.5281/zenodo.5137552>. Acesso: 06 jun. 2022.

VIEIRA, André; PRÍNCIPE, Pedro. Contributos do projeto FAIRsFAIR para a implementação das políticas de dados FAIR: dos repositórios aos currículos académicos. *In*: CONFERÊNCIA LUSO-BRASILEIRA DE CIÊNCIA ABERTA, 11., Braga, Portugal, 2021. [Realizada de 06 a 08 de outubro de 2020, em formato virtual]. Disponível em: <https://hdl.handle.net/1822/69067>. Acesso em: 05 jun. 2022.

WILKINSON, Mark D. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. **Sci Data**, v. 3, n. 160018, 2016. Disponível em: <https://doi.org/10.1038/sdata.2016.18>. Acesso em: 06 jun. 2022.

ZIMAN, John. **Conhecimento Público**. Tradução Regina Reis Junqueira. Belo Horizonte: Itatiaia; São Paulo: EDUSP, 1979. 164 p. (Coleção O homem e a Ciência, v. 8)

## APÊNDICE A – CICLO DE VIDA DATAONE

FASE DO PLANO	
Ação	Questões/Ações/Recomendações
Colete os seus dados	Com base nas hipóteses e no plano de amostragem, quais dados serão gerados?
	Como as amostras serão coletadas e analisadas?
	Observação: forneça documentação descritiva do raciocínio, dos métodos de coleta, dos métodos de análise e qualquer informação contextual relevante.
	Quais instrumentos serão usados?
Decida sobre um repositório	Que infraestrutura cibernética será necessária para apoiar a sua pesquisa?
	Selecione um repositório de dados (ou seja, centro de dados) que seja mais apropriado para os dados que você irá gerar e para a comunidade que fará uso deles.
	Converse com colegas e patrocinadores de pesquisas sobre o melhor repositório para a sua disciplina e o seu tipo de dado.
	Verifique com o repositório sobre os requisitos para envio, incluindo documentação necessária dos dados, padrões de metadados e quaisquer possíveis restrições de uso (por exemplo, direitos de propriedade intelectual).
Organize os seus dados	As especificações do repositório ajudarão a orientar as decisões sobre o restante das práticas abaixo.
	Decida como os dados serão organizados em um arquivo, quais formatos de arquivo serão usados e os tipos de produtos de dados que serão gerados.
	Sua comunidade tem formatos padrão (formatos de arquivo, unidades, nomes de parâmetros)?
Gerencie os seus dados	Observação: considere se um banco de dados relacional ou outra estratégia de organização de dados pode ser mais apropriado para sua pesquisa.
	Quem é responsável por gerenciar os dados?
	Como será o controle de versão manipulado?
Descreva os seus dados	Como será feito o <i>backup</i> dos dados e com que frequência?
	Como você produzirá um registro de metadados?
	Qual padrões de metadados você utilizará?
	Usando qual ferramenta?
	Você criará um registro no início do projeto e o atualizará à medida que avança em sua pesquisa?
	Onde você vai depositar os metadados?
Compartilhe os seus dados	Observação: considere os padrões de sua comunidade ao decidir sobre o padrão de metadados e o <i>data center</i> .
	Desenvolva um plano para compartilhar dados com a equipe do projeto, com outros colaboradores e com a comunidade científica mais ampla.
	Em que condições os dados serão divulgados para cada um desses grupos?
	Quais são as datas de lançamento para esses grupos?
Preserve os seus dados	Como os dados serão divulgados?
	À medida que os arquivos são criados, implemente um plano de preservação de dados de curto prazo que garanta que os dados possam ser recuperados em caso de perda de arquivos.
	Por exemplo, armazenar dados rotineiramente em locais diferentes.

Considere o seu orçamento	Que tipo de pessoal será necessário para realizar um plano de gerenciamento para os seus dados?
	Que tipos de <i>hardwares</i> , <i>softwares</i> ou outros recursos computacionais serão necessários?
	Que outras despesas podem ser necessárias, como doação ou pagamento do <i>data center</i> ? Observação: o orçamento preparado para o seu projeto de pesquisa deve incluir custos estimados para o gerenciamento de dados.
Explore os seus recursos institucionais	Algumas instituições têm modelos de planos de gerenciamento de dados, <i>data centers</i> institucionais sugeridos, sugestões de orçamento e ferramentas úteis para planejar seu projeto.
<b>FASE DA COLETA</b>	
Descreva o conteúdo de seus arquivos de dados	Defina cada parâmetro, incluindo seu formato, as unidades usadas e os códigos para valores ausentes.
	Forneça exemplos de formatos para parâmetros comuns.
	As descrições de dados devem acompanhar os arquivos de dados como um arquivo “ <i>readme.txt</i> ”, um arquivo de metadados deve usar um padrão de metadados aceito.  Recomendação: organize os dados em um arquivo de uma das duas maneiras descritas abaixo. Seja qual for o estilo que você usar, certifique-se de colocar cada observação em uma linha separada.
Uso de dados consistente da organização	Recomendamos que você organize os dados em um arquivo de uma das duas maneiras descritas. Seja qual for o estilo que você usar, certifique-se de colocar cada observação em uma linha separada: <ol style="list-style-type: none"> <li>1. Cada linha de um arquivo representa um registro completo, já as colunas representam todos os parâmetros que compõem o registro (formato de planilha);</li> <li>2. Uma coluna é usada para definir o parâmetro e outra coluna é usada para o valor do parâmetro (um formato de banco de dados). Outras colunas podem ser usadas para local, data, tratamento, unidades de medida etc.</li> </ol>
	Use o mesmo formato em todo o arquivo; por exemplo, não reorganize colunas ou linhas.
	Na parte superior do arquivo, inclua uma ou mais linhas de cabeçalho que identifiquem o parâmetro e as unidades de cada coluna. “Atomize” dados: certifique-se de que haja apenas um dado em cada entrada.
Caracteres de texto	Use caracteres ASCII de texto simples para nomes de variáveis, nomes de arquivos e dados. Comentário: isso garantirá que seu arquivo de dados seja legível pelo número máximo de programas de <i>software</i> .
Use <i>software</i> e <i>hardware</i> estáveis e não proprietários	Os formatos de arquivo devem, idealmente, ser não proprietários (por exemplo, arquivos <i>.txt</i> ou <i>.csv</i> em vez de <i>.xls</i> ), para que sejam estáveis e possam ser lidos no futuro.
	Considere a longevidade do <i>hardware</i> ao fazer um <i>backup</i> de dados.
Atribua nomes aos arquivos descritivos	Os nomes dos arquivos descrevem idealmente o projeto, o conteúdo do arquivo, o local e a data, portanto, devem ser exclusivos o suficiente para serem independentes como descrições de arquivo.
	Os nomes dos arquivos não substituem os registros de metadados completos.
Mantenha seus dados brutos	Preserve os dados brutos, com todas as suas imperfeições.
	Use um <i>script</i> programa para “limpar” os dados, assim, todas as etapas do processo serão documentadas.

Crie uma tabela de parâmetros	Descreva o código e as abreviações usadas para um parâmetro, as unidades, os valores máximos e mínimos, o tipo de dado (ou seja, texto, numérico) e uma descrição.
Crie uma tabela de sites	Descreva os locais onde os dados foram coletados, incluindo latitude, longitude, datas visitadas e quaisquer detalhes contextuais que podem afetar a coleta.  Por exemplo, tipo de ecossistema, cobertura ou uso da terra, condições climáticas etc.
<b>FASE DE ASSEGURAR A QUALIDADE</b>	
Garantia básica de qualidade e de controle de qualidade dos dados	Descreva quaisquer condições durante a coleta que possam afetar a qualidade dos dados.
	Identifique os valores estimados, verifique novamente os dados inseridos manualmente (de preferência inseridos por mais de uma pessoa) e use sinalizadores de nível de qualidade para indicar possíveis problemas.
	Verifique o formato dos dados para ter certeza de que é consistente em todo o conjunto.
	Realize resumos estatísticos e gráficos (por exemplo, máx./mín., média, intervalo) para verificar valores questionáveis ou impossíveis e identificar discrepâncias.
Comunique a qualidade dos dados	Use a codificação no conjunto de dados que indica a qualidade ou nos metadados ou na documentação de dados.
	Identifique os valores ausentes.
	Verifique os dados usando os conjuntos de dados semelhantes para identificar possíveis problemas. <b>Observação:</b> problemas adicionais com os dados também podem ser identificados durante a análise e a interpretação, antes da preparação do manuscrito.
<b>FASE DE DESCREVER</b>	
Os metadados devem ser gerados em um formato comumente utilizado pela comunidade científica mais relevante.	
Descreva o contexto digital	Nome do conjunto de dados.
	O(s) nome(s) do(s) arquivo(s) de dados no conjunto de dados.
	Data em que o conjunto de dados foi modificado pela última vez.
	Registros de arquivo de dados de exemplo para cada arquivo de tipo de dados.
	Arquivos complementares pertinentes.
	Lista de conjuntos de dados relacionados ou auxiliares.
	<i>Software</i> (incluindo número de versão) usado para preparar/ler o conjunto de dados.
Data que o processamento de dados foi realizado.	
Descreva o pessoal e as partes interessadas	Quem coletou os dados.
	Quem deve ser contatado para perguntas.
	Quem são os patrocinadores.
Descreva o contexto científico	Razão científica pela qual os dados foram coletados.

	Quais dados foram coletados.
	Quais instrumentos (incluindo modelo e número de série) foram usados.
	Condições ambientais durante a coleta.
	Onde foram coletados e qual a resolução espacial.
	Quando foram coletados e qual a resolução temporal.
	Padrões ou calibrações usadas.
Descreva informações sobre os parâmetros	Como cada um foi medido ou produzido.
	Unidades de medida.
	Formato usado no conjunto de dados.
	Precisão, exatidão e incerteza.
	Informações sobre dados.
	Detalhes taxonômicos.
	Definições de códigos usados.
	Garantia de qualidade e de atividades.
	Problemas conhecidos que limitam o uso de dados. Exemplo: incertezas e problemas de amostragem.
Como citar o conjunto de dados.	
<b>FASE DE PRESERVAR</b>	
Identifique dados com valor a longo prazo	Não é necessário arquivar todos os produtos de dados gerados a partir de sua pesquisa.
	Considere o tamanho de seus arquivos, quais dados serão mais úteis para futuros usuários (normalmente os dados brutos) e quais versões de dados seriam mais difíceis de reproduzir.
Armazene dados usando precisão apropriada	Por exemplo, dígitos significativos.
Use terminologia padrão	Para permitir que outras pessoas encontrem e usem seus dados, selecione, cuidadosamente, que terminologia irá descrevê-los.
	Use palavras-chave comuns e consulte ontologias para sua disciplina, se estiverem disponíveis.
Considere políticas legais e outras	Toda pesquisa requer o compartilhamento de informações e de dados.
	1. Verifique as políticas de privacidade e confidencialidade de sua instituição;
	2. Os dados não são protegidos por direitos autorais. Certifique-se de ter as permissões apropriadas quando usando dados que tenham vários proprietários ou camadas de direitos autorais. As informações que documentam o contexto da coleta de dados podem estar sob direitos;
	3. Os dados podem ser licenciados. A maneira pela qual você licencia seus dados pode determinar sua capacidade de ser consumido por outros acadêmicos. Por exemplo, a Licença <i>Creative Commons Zero</i> [14] prevê um acesso muito amplo.
Se os seus dados se enquadrarem em qualquer uma das categorias, há	Espécies raras, ameaçadas ou ameaçadas de extinção;
	Itens culturais devolvidos ao país de origem;

considerações adicionais em relação ao compartilhamento	Restos humanos e objetos nativos americanos e nativos havaianos;
	Qualquer pesquisa envolvendo seres humanos;
	Se você usar dados de outras fontes, revise seus direitos de uso dos dados e certifique-se de ter as licenças e as permissões apropriadas.
Atribuição e proveniência	O pessoal responsável pelo conjunto de dados durante toda a vida útil do conjunto de dados.
	O contexto do conjunto de dados em relação a um projeto ou estudo maior (incluindo <i>links</i> e documentação relacionada), se aplicável.
	Histórico de revisões, incluindo adições de novos dados e correções de erros.
	<i>Links</i> para dados de origem, se os dados foram derivados de outro conjunto de dados.
	Suporte ao projeto (por exemplo, agências de financiamento, colaboradores e suporte computacional).
	Como citar corretamente o conjunto de dados.
	Direitos de propriedade intelectual e outras considerações de licenciamento.
<b>FASES DE DESCOBRIR, INTEGRAR E ANALISAR</b>	
Uma variedade de ferramentas está disponível para dar suporte à descoberta, à integração, à análise e à visualização de dados.	
Avanços recentes significativos foram feitos no suporte à criação e ao gerenciamento de fluxos de trabalho científicos complexos que servem para integrar, analisar e visualizar dados, bem como, documentar as etapas exatas usadas nesses processos.	
Proveniência do conjunto de dados	Quando conjuntos de dados e elementos de dados são usados como fonte para novos conjuntos, é importante identificar e documentar esses dados na documentação do conjunto derivado.
	1) rastrear o uso de conjuntos de dados e elementos de dados; 2) atribuir aos criadores dos conjuntos de dados originais; 3) identificar efeitos de erros nos conjuntos de dados originais em conjuntos de dados derivados.

Fonte: Strasser *et al.* (2012).



## APÊNDICE B – CRITÉRIOS CORE TRUST SEAL

### 0. Informações básicas/Contexto

#### R0 – Forneça contexto para o seu repositório.

- Tipo de repositório: esse item ajudará os revisores a entenderem qual função seu repositório executa.
- Breve descrição do repositório.
- Breve descrição da comunidade designada do repositório.
- Nível de curadoria realizada:
  - conteúdo distribuído conforme depositado;
  - curadoria básica – por exemplo, breve certificação, adição de metadados básicos ou documentação;
  - curadoria aprimorada – por exemplo, conversão para novos formatos e aprimoramento da documentação;
  - curadoria ao nível de dados – como na aprimorada, mas com edição adicional de dados depositados para precisão.
- Parceiros terceirizados: se eles existem, liste-os.
- Outras informações relevantes: o repositório pode querer adicionar informações contextuais extras que não são abordadas nos requisitos, mas que podem ser úteis para os revisores fazerem sua avaliação. Por exemplo, você pode descrever:
  - o uso e o impacto dos acervos de dados do repositório (citações, uso por outros projetos etc.);
  - uma função nacional, regional ou global que o repositório atende;
  - qualquer *cluster* global ou organização de rede à qual o repositório pertença.

### INFRAESTRUTURA ORGANIZACIONAL

#### 1. Missão/Esopo

##### R1 – O repositório tem a missão explícita de fornecer acesso e preservar os dados em seu domínio.

Para esse requisito, descreva:

- a missão da sua organização em preservar e fornecer acesso a dados, inclua *links* para declarações explícitas deste objetivo;
- o nível de aprovação que a missão recebeu dentro da organização.

A evidência para este requisito pode assumir a forma de uma declaração de missão pública aprovada, funções exigidas pelos financiadores ou declaração de política assinada pelo conselho de administração.

#### 2. Licenças

##### R2 – O repositório mantém todas as licenças que abrangem o acesso e uso de dados e monitora a conformidade.

– Os repositórios devem manter todas as licenças aplicáveis que abrangem o acesso e o uso de dados, deve, inclusive, comunicar sobre elas aos usuários e monitorar a conformidade.

Para esse requisito, descreva:

- contratos de licença em uso;
- condições de uso (direitos de propriedade intelectual, distribuição, uso pretendido, proteção de dados sensíveis etc.);
- documentação sobre medidas em caso de descumprimento das condições de acesso e de uso.

**Observação:** as disposições éticas e de privacidade que afetam as licenças são tratadas em R4 (Confidencialidade/Ética). A garantia de que as licenças de depósito fornecem direitos suficientes para o repositório manter, preservar e oferecer acesso aos dados deve ser coberta pelo R10 (Plano de preservação).

### 3. Continuidade de acesso

**R3 – O repositório tem um plano de continuidade para garantir o acesso contínuo e a preservação do seu acervo.**

– Esse requisito abrange a governança relacionada à operação continuada do repositório ao longo do tempo e durante desastres, bem como, as evidências em relação ao planejamento sucessório, ou seja, as medidas em vigor para garantir o acesso e a disponibilidade dos acervos de dados, tanto no presente como no futuro.

Para esse requisito, descreva:

- o nível de responsabilidade assumido pelos acervos de dados, incluindo qualquer garantida de períodos de conservação;
- os planos de médio prazo (de três a cinco anos) e de longo prazo (> que cinco anos) em vigor para garantir a disponibilidade e a acessibilidade contínuas dos dados.

**Observação:** as evidências para esse requisito devem estar relacionadas especificamente à governança. Os aspectos técnicos de continuidade de negócios e de planejamento de desastres e sucessão devem ser abordados em R15 (Infraestrutura técnica).

### 4. Confidencialidade/Ética

**R4 – O repositório garante, na medida do possível, que os dados sejam criados, curados, acessados e usados em conformidade com as normas disciplinares e éticas.**

Para esse requisito, as respostas devem incluir evidências relacionadas às seguintes perguntas:

- Como o repositório cumpre as normas disciplinares aplicáveis?
- O repositório solicita confirmação de que a coleta ou criação de dados foi realizada de acordo com os critérios legais e éticos prevalecentes na localização geográfica ou disciplina do produtor de dados (por exemplo, Comitê de Revisão Ética/Conselho de Revisão Institucional ou Legislação de Proteção de Dados)?
- São aplicados procedimentos especiais para gerenciar os dados com risco de divulgação?
- Os dados com risco de divulgação são armazenados adequadamente para limitar o acesso?
- Os dados com risco de divulgação são distribuídos em condições apropriadas?
- Existem procedimentos em vigor para revisar o risco de divulgação nos dados e tomar as medidas necessárias para anonimizar arquivos ou fornecer acesso de forma segura?
- Os funcionários são treinados no gerenciamento de dados com risco de divulgação?
- Existem medidas em vigor se as condições não forem cumpridas?
- O repositório fornece orientação sobre o uso responsável de informações divulgadas ou potencialmente divulgadas?

**Observação:** esse requisito é sobre as disposições éticas e de privacidade que afetam a criação, a curadoria e o uso dos dados. Detalhes sobre licenças em alinhamento com tais disposições éticas e de privacidade devem ser cobertos em R2 (Licenças).

### 5. Infraestrutura organizacional

**R5 – O repositório tem financiamento adequado e um número suficiente de pessoal qualificado gerido através de um sistema claro de governação para levar a cabo a missão de forma eficaz.**

Para esse requisito, as respostas devem incluir evidências relacionadas aos seguintes assuntos:

- o repositório é hospedado por uma instituição reconhecida (garantindo estabilidade e sustentabilidade de longo prazo) apropriada para sua Comunidade Designada;
- o repositório tem financiamento suficiente, incluindo recursos de equipe, recursos de TI e um orçamento para participar de reuniões quando necessário. Idealmente, isso deve ser por um período de três a cinco anos;
- o repositório garante que seus funcionários tenham acesso a treinamento contínuo e desenvolvimento constante;
- o alcance e a profundidade da especialização, tanto da organização quanto de sua equipe, incluindo quaisquer afiliações relevantes (por exemplo, órgãos nacionais ou internacionais), são apropriados para a missão.

**Observação:** o acesso a aconselhamento especializado objetivo, além do fornecido por pessoal qualificado, é coberto em R6 (Orientação especializada).

### 6. Orientação especializada

**R6 – O repositório adota mecanismo(s) para garantir orientação e *feedback* contínuos de especialistas (internos ou externos, incluindo orientação científica, se relevante).**

Para esse requisito, as respostas devem incluir evidências relacionadas às seguintes perguntas:

- O repositório possui consultores internos ou um comitê consultivo externo que pode ser composto por especialistas técnicos, de curadoria, de ciência de dados e de disciplinares?
- Como o repositório se comunica com os especialistas para aconselhamento?
- Como o repositório se comunica com sua comunidade designada para *feedback*?

**Observação:** esse requisito procura confirmar que o repositório tem acesso ao aconselhamento especializado objetivo, além do fornecido pelo pessoal qualificado mencionado em R5 (Infraestrutura organizacional).

## GERENCIAMENTO DE OBJETOS DIGITAIS

### 7. Integridade e autenticidade dos dados

**R7 – O repositório garante a integridade e a autenticidade dos dados.**

Para esse requisito, as respostas sobre integridade de dados devem incluir evidências relacionadas ao seguinte:

- descrição das verificações para ver se um objeto digital não foi alterado ou corrompido (ou seja, checagens fixas) do depósito ao uso;
- documentação da completude dos dados e metadados;
- detalhes de como todas as alterações nos dados e metadados são registradas;
- descrição da estratégia de controle de versão;
- uso de padrões e de convenções internacionais apropriados (que devem ser especificados).

A evidência de gerenciamento de autenticidade deve estar relacionada às seguintes perguntas:

- O repositório possui uma estratégia para alterações de dados? Os produtores de dados estão cientes disso?
- O repositório mantém dados de proveniência e de trilhas de auditoria relacionadas?
- O repositório mantém *links* para metadados e outros conjuntos de dados? Se sim, como?
- O repositório compara as propriedades essenciais de diferentes versões do mesmo arquivo? Como?
- O repositório verifica as identidades dos depositantes?

**8. Avaliação****R8 – O repositório aceita dados e metadados com base em critérios definidos para garantir relevância e compreensão para os usuários dos dados.**

Para esse requisito, as respostas devem incluir evidências relacionadas às seguintes perguntas:

- O repositório usa uma política de desenvolvimento de coleção para orientar a seleção de dados para arquivamento?
- Qual abordagem é usada para dados que não se enquadram no perfil de missão/coleção?
- O repositório possui procedimentos para determinar se os metadados necessários para interpretar e usar os dados são fornecidos?
- Existe alguma avaliação automatizada de aderência de metadados a esquemas relevantes?
- Qual é a abordagem do repositório se os metadados fornecidos forem insuficientes para longo prazo de preservação?
- O repositório publica uma lista de formatos preferidos?
- Qual é a abordagem para dados que são depositados em formatos não preferenciais?
- Qual é o processo para remover itens de sua coleção, levando em conta também o impacto dos identificadores persistentes existentes?

**Observação:** esse requisito abrange os critérios de seleção aplicados no ponto de depósito. A qualidade dos dados e a melhoria durante o processo de curadoria devem ser cobertas pelo R11 (Qualidade dos dados).

**9. Procedimentos de armazenamento documentados****R9 – O repositório aplica processos e procedimentos documentados no gerenciamento do armazenamento de dados em arquivo.**

Para esse requisito, as respostas devem incluir evidências relacionadas às seguintes perguntas:

- Como os processos e os procedimentos relevantes são documentados e gerenciados?
- O repositório tem uma compreensão clara de todos os locais de armazenamento e como eles são gerenciados?
- O repositório tem uma estratégia para várias cópias? Se assim for, qual é?
- As técnicas de gerenciamento de risco são usadas para informar a estratégia?
- Quais verificações estão em vigor para garantir a consistência entre as cópias de arquivo?
- Como a deterioração da mídia de armazenamento é tratada e monitorada?

**Observação:** detalhes sobre a implementação técnica de armazenamento devem ser cobertos em R15 (Infraestrutura técnica) e arranjos específicos para segurança física e lógica em R16 (Segurança).

**10. Plano de preservação**

**R10 – O repositório assume a responsabilidade pela preservação a longo prazo e gere essa função de forma planejada e documentada.**

Para esse requisito, as respostas devem incluir evidências relacionadas às seguintes perguntas:

- O repositório tem uma abordagem documentada para preservação?
- O nível de responsabilidade pela preservação de cada item é compreendido? Como isso é definido?
- Existem planos relacionados a futuras migrações ou medidas semelhantes para lidar com a ameaça de obsolescência?
- O contrato entre depositante e depositário prevê todas as ações necessárias para cumprir as responsabilidades?
- A transferência de custódia e responsabilidade está clara para o depositante e o depositário?
- O repositório tem direitos para copiar, transformar e armazenar os itens, bem como, fornecer acesso a eles?
- As ações relevantes para preservação estão especificadas na documentação, incluindo transferência de custódia, padrões de informações de submissão e padrões de informações de arquivamento?
- Existem medidas para garantir que essas ações sejam tomadas?

**Observação:** os direitos relativos ao acesso e ao uso de dados e o monitoramento de sua conformidade devem ser cobertos pelo R2 (Licenças).

**11. Qualidade dos dados**

**R11 – O repositório possui conhecimento adequado para tratar da qualidade de dados técnicos e metadados e garante que informações suficientes estejam disponíveis para que os usuários finais façam avaliações relacionadas à qualidade.**

Para esse requisito, descreva:

- a abordagem de qualidade de dados e metadados adotada pelo repositório. O repositório possui verificações de controle de qualidade para garantir a integridade e a compreensibilidade dos dados depositados? Em caso afirmativo, forneça referências a padrões de controle de qualidade e mecanismos de relatórios aceitos pela comunidade de prática relevante e inclua detalhes de como quaisquer problemas são resolvidos (por exemplo, os dados são devolvidos ao provedor de dados para retificação, corrigidos pelo repositório, anotados por sinalizadores de qualidade no arquivo de dados e/ou incluídos nos metadados que o acompanham?);
- a capacidade da comunidade designada de comentar e/ou classificar dados e metadados;
- se são fornecidas citações de trabalhos relacionados ou *links* para índices de citações.

**Observação:** Esse requisito refere-se a padrões de qualidade de dados e à garantia durante a curadoria. Os critérios de seleção são abordados em R8 (Avaliação).

### 12. Fluxos de trabalho

**R12 – O arquivamento ocorre de acordo com fluxos de trabalho definidos, desde a ingestão até a disseminação.**

Para esse requisito, as respostas devem incluir evidências relacionadas aos seguintes pontos:

- as descrições de fluxos de trabalho/processos de negócios;
- a comunicação clara aos depositantes e aos usuários sobre o manuseio de dados;
- os níveis de segurança e os impactos nos fluxos de trabalho (proteção da privacidade dos assuntos etc.);
- a verificação qualitativa e quantitativa dos resultados;
- a avaliação e a seleção de dados;
- as abordagens para dados que não se enquadram no perfil de missão/coleção;
- os tipos de dados gerenciados e qualquer impacto no fluxo de trabalho;
- a manipulação de decisões nos fluxos de trabalho (por exemplo, transformação de dados de arquivamento);
- o gerenciamento de mudanças de fluxos de trabalho.

**Observação:** Esse requisito confirma que todos os fluxos de trabalho estão documentados.

### 13. Descoberta e identificação de dados

**R13 – O repositório permite que os usuários descubram os dados e os consultem de forma persistente por meio da citação adequada.**

Para esse requisito, as respostas devem incluir evidências relacionadas às seguintes perguntas:

- O repositório oferece recursos de pesquisa?
- O repositório mantém um catálogo de metadados pesquisáveis de acordo com os padrões apropriados (acordados internacionalmente)?
- Quais sistemas de identificadores persistentes o repositório usa?
- O repositório facilita a coleta automática dos metadados?
- O repositório está incluído em um ou mais registros disciplinares ou genéricos de recursos?
- O repositório oferece citações de dados recomendadas?

### 14. Reutilização de dados

**R14 – O repositório permite a reutilização dos dados ao longo do tempo, garantindo que os metadados apropriados estejam disponíveis para apoiar a compreensão e o uso dos dados.**

Para esse requisito, as respostas devem incluir evidências relacionadas às seguintes perguntas:

- Quais metadados são fornecidos pelo repositório quando os dados são acessados?
- Como o repositório garante a compreensão contínua dos dados?
- Os dados são fornecidos em formatos usados pela comunidade designada? Quais formatos?
- São tomadas medidas para levar em conta a possível evolução dos formatos?

O conceito de “reutilização” é crítico em ambientes nos quais os resultados da análise secundária são depositados novamente em um repositório junto com os dados primários, uma vez que, a cadeia de proveniências e as questões de direitos associadas podem se tornar cada vez mais complicadas.

## TECNOLOGIA

### 15. Infraestrutura técnica

**R15 – O repositório funciona em sistemas operacionais bem suportados e outro *software* de infraestrutura central e está usando tecnologias de *hardware* e *software* apropriadas aos serviços que fornece à sua comunidade designada.**

Os repositórios precisam operar em infraestruturas centrais confiáveis e estáveis que maximizem a disponibilidade do serviço. Além disso, o *hardware* e o *software* usados devem ser relevantes e apropriados para a comunidade designada e para as funções que um repositório cumpre. O modelo de referência OAIS especifica as funções de um repositório para atender às necessidades do usuário.

Para esse requisito, as respostas devem incluir evidências relacionadas às seguintes perguntas:

- Quais padrões o repositório usa para referência? Esses padrões são internacionais e/ou comunitários? Com que frequência eles são revisados?
- Como os padrões são implementados? Existem desvios significativos de padrão? Se sim, por favor explique.

- O repositório tem um plano para desenvolvimento de infraestrutura? Se assim for, qual é?
- Um inventário de *software* é mantido e a documentação do sistema está disponível?
- O *software* suportado pela comunidade está em uso? Por favor, descreva.
- A disponibilidade, a largura de banda e a conectividade são suficientes para atender às necessidades da comunidade?
- O repositório possui um plano de desastre e um plano de continuidade de negócios? Em particular, existem procedimentos e arranjos para fornecer recuperação rápida ou *backup* de serviços essenciais em caso de interrupção? Quais são eles?

**Observação:** os aspectos de governança de continuidade de negócios, de planejamento de desastres e de planejamento de sucessão devem ser abordados em R3 (Continuidade de acesso). Os detalhes sobre o processo de armazenamento devem ser abordados em R9 (Procedimentos de armazenamento documentados). Os arranjos de segurança são cobertos em R16 (Segurança).

## 16. Segurança

**R16 – A infraestrutura técnica do repositório fornece proteção da instalação e seus dados, produtos, serviços e usuários.**

O repositório deve analisar ameaças potenciais, avaliar riscos e criar um sistema de segurança consistente. Ele deve descrever cenários de danos com base em ações maliciosas, erros humanos ou falhas técnicas que representam uma ameaça ao repositório e aos seus dados, produtos, serviços e usuários. Ele deve medir a probabilidade e o impacto de tais cenários, decidir quais níveis de risco são aceitáveis e determinar quais medidas devem ser tomadas para combater as ameaças ao repositório e a sua comunidade designada. Esse deve ser um processo contínuo.

Para esse requisito, descreva:

- seu sistema de segurança de TI, seus funcionários com funções relacionadas à segurança (por exemplo, agentes de segurança) e quaisquer ferramentas de análise de risco (por exemplo, DRAMBORA1) que você usa;
- quais os níveis de segurança são necessários e como eles são suportados;
- Quaisquer são os procedimentos de autenticação e de autorização empregados para gerenciar com segurança o acesso aos sistemas em uso (por exemplo, Shibboleth, OpenAthens).

**Observação:** os processos de armazenamento e de infraestrutura técnica que utilizam essas medidas de segurança devem ser abordados em R9 (Procedimentos de armazenamento documentado) e R15 (Infraestrutura Técnica), respectivamente.

## APÊNDICE C – CAMPOS DO DMPONLINE

<b>Coleção de dados</b>		
	Questões a considerar	Orientação
<b>Quais dados você coletará ou criará?</b>	<ul style="list-style-type: none"> <li>– Qual o tipo, o formato e o volume dos dados?</li> <li>– Os formatos e <i>softwares</i> escolhidos permitem o compartilhamento e o acesso de longo prazo aos dados?</li> <li>– Existem dados existentes que você pode reutilizar?</li> </ul>	Forneça uma breve descrição dos dados, incluindo quaisquer dados existentes ou fontes de terceiros que serão usados, observando em cada caso seu conteúdo, tipo e cobertura. Descreva e justifique sua escolha de formato e considere as implicações do formato de dados e volumes de dados em termos de armazenamento, <i>backup</i> e acesso.
<b>Como os dados serão coletados ou criados?</b>	<ul style="list-style-type: none"> <li>– Quais padrões ou metodologias você usará?</li> <li>– Como você estruturará e nomeará suas pastas e seus arquivos?</li> <li>– Como você vai lidar com o versionamento?</li> <li>– Quais processos de garantia de qualidade você adotará?</li> </ul>	Descreva como os dados serão coletados/criados e quais padrões de dados da comunidade (se houver) serão usados. Considere como os dados serão organizados durante o projeto, mencionando, por exemplo, convenções de nomenclatura, controle de versão e estruturas de pastas. Explique como a consistência e a qualidade da coleta de dados serão controladas e documentadas. Isso pode incluir processos como calibração, amostras ou medições repetidas, captura ou registro de dados padronizados, validação de entrada de dados, revisão por pares de dados ou representação com vocabulários controlados.
<b>Documentação e metadados</b>		
<b>Que documentação e metadados acompanharão os dados?</b>	<ul style="list-style-type: none"> <li>– Quais informações são necessárias para que os dados sejam lidos e interpretados no futuro?</li> <li>– Como você irá capturar/criar esta documentação e metadados?</li> <li>– Quais padrões de metadados você usará e por quê?</li> </ul>	Descreva os tipos de documentação que acompanharão os dados para ajudar os usuários secundários a entendê-los e reutilizá-los. Isso deve incluir pelo menos detalhes básicos que ajudarão as pessoas a encontrar os dados, incluindo quem criou ou contribuiu, seu título, data de criação e sob quais condições eles podem ser acessados. A documentação também pode incluir detalhes sobre a metodologia utilizada, informações analíticas e procedimentais, definições de variáveis, vocabulários, unidades de medida, quaisquer suposições feitas e o formato e tipo de arquivo dos dados. Considere como você capturará essas informações e onde elas serão registradas. Sempre que possível, identifique e use os padrões comunitários existentes.
<b>Ética e conformidade legal</b>		
<b>Como você vai gerenciar quaisquer questões éticas?</b>	<ul style="list-style-type: none"> <li>– Você obteve consentimento para a preservação e o compartilhamento de dados?</li> <li>– Como você protegerá a identidade dos participantes, se necessário? Por exemplo, através de anonimização.</li> </ul>	As questões éticas afetam como você armazena os dados, quem pode vê-los/usá-los e por quanto tempo eles são mantidos. A gestão de preocupações éticas pode incluir: anonimização de dados; encaminhamento a comitês de ética departamentais ou institucionais;



	<ul style="list-style-type: none"> <li>– Como os dados confidenciais serão tratados para garantir que sejam armazenados e transferidos com segurança?</li> </ul>	e acordos formais de consentimento. Você deve mostrar que está ciente de quaisquer problemas e que se planejou adequadamente. Se você estiver realizando pesquisas envolvendo participantes humanos, também deve garantir que o consentimento seja solicitado para permitir que os dados sejam compartilhados e reutilizados.
<p><b>Como você gerenciará questões de direitos autorais e direitos de propriedade intelectual (DPI)?</b></p>	<ul style="list-style-type: none"> <li>– Quem é o proprietário dos dados?</li> <li>– Como os dados serão licenciados para reutilização?</li> <li>– Existem restrições à reutilização de dados de terceiros?</li> <li>– O compartilhamento de dados será adiado/restringido, por exemplo, para publicar ou buscar patentes?</li> </ul>	Declare quem será o proprietário dos direitos autorais e dos DPI de quaisquer dados que você coletar ou criar, juntamente com a(s) licença(s) para seu uso e reutilização. Para projetos com vários parceiros, pode valer a pena cobrir a propriedade de DPI em um contrato de consórcio. Considere quaisquer políticas relevantes de financiadores, instituições, departamentos ou grupos sobre direitos autorais ou DPI. Considere também as permissões para reutilizar dados de terceiros e quaisquer restrições necessárias ao compartilhamento de dados.
<b>Armazenamento e backup</b>		
<p><b>Como os dados serão armazenados e copiados durante a pesquisa?</b></p>	<ul style="list-style-type: none"> <li>– Você tem armazenamento suficiente ou precisará incluir cobranças por serviços adicionais?</li> <li>– Como será feito o <i>backup</i> dos dados?</li> <li>– Quem será responsável pelo <i>backup</i> e pela recuperação?</li> <li>– Como os dados serão recuperados no caso de um incidente?</li> </ul>	Indique com que frequência os dados serão copiados e em quais locais. Quantas cópias estão sendo feitas? Armazenar dados apenas em <i>laptops</i> , discos rígidos de computadores ou dispositivos de armazenamento externos é muito arriscado. O uso de armazenamento robusto e gerenciado fornecido por equipes de TI da universidade é preferível. Da mesma forma, normalmente é melhor usar serviços de <i>backup</i> automáticos fornecidos por serviços de TI do que confiar em processos manuais. Se você optar por usar um serviço de terceiros, garanta que isso não entre em conflito com nenhuma política do financiador, institucional, departamental ou de grupo, por exemplo, em termos da jurisdição legal em que os dados são mantidos ou a proteção de dados confidenciais.
<p><b>Como você gerenciará o acesso e a segurança?</b></p>	<ul style="list-style-type: none"> <li>– Quais são os riscos para a segurança dos dados e como eles serão gerenciados?</li> <li>– Como você controlará o acesso para manter os dados seguros?</li> <li>– Como você garantirá que os colaboradores possam acessar seus dados com segurança?</li> <li>– Se estiver criando ou coletando dados em campo, como você garantirá uma transferência segura para seus principais sistemas protegidos?</li> </ul>	Se seus dados são confidenciais (por exemplo, dados pessoais que ainda não são de domínio público, informações confidenciais ou segredos comerciais), você deve descrever quaisquer medidas de segurança apropriadas e observar quaisquer padrões formais que você cumprirá, por exemplo, ISO 27001.
<b>Seleção e preservação</b>		
<p><b>Quais dados são de valor a longo prazo e devem ser retidos,</b></p>	<ul style="list-style-type: none"> <li>– Quais dados devem ser retidos/destruídos para fins contratuais, legais ou regulatórios?</li> <li>– Como você decidirá quais dados manter?</li> </ul>	Considere como os dados podem ser reutilizados, por exemplo, para validar seus resultados de pesquisa, conduzir novos estudos ou para ensinar. Decida quais dados manter e por quanto tempo. Isso pode se

<b>compartilhados e/ou preservados?</b>	<ul style="list-style-type: none"> <li>– Quais são os usos previsíveis de pesquisa para os dados?</li> <li>– Por quanto tempo os dados serão retidos e preservados?</li> </ul>	basear em quaisquer obrigações de reter determinados dados, o valor potencial de reutilização, o que é economicamente viável manter e qualquer esforço adicional necessário para preparar os dados para compartilhamento e preservação de dados. Lembre-se de considerar qualquer esforço adicional necessário para preparar os dados para compartilhamento e preservação, como alterar formatos de arquivo.
<b>Qual é o plano de preservação a longo prazo para o conjunto de dados?</b>	<ul style="list-style-type: none"> <li>– Onde, por exemplo, em qual repositório ou arquivo os dados serão mantidos?</li> <li>– Quais custos, se houver, serão cobrados pelo repositório ou arquivo de dados selecionado?</li> <li>– Você gastou tempo e esforço para preparar os dados para compartilhamento/preservação?</li> </ul>	Considere como os conjuntos de dados que têm valor de longo prazo serão preservados e curados além da duração da concessão. Também descreva os planos para preparar e documentar dados para compartilhamento e arquivamento. Se você não propõe o uso de um repositório estabelecido, o plano de gerenciamento de dados deve demonstrar que os recursos e sistemas estarão disponíveis para permitir que os dados sejam curados de forma eficaz além da duração da concessão.
<b>Compartilhamento de dados</b>		
<b>Como você compartilhará os dados?</b>	<ul style="list-style-type: none"> <li>– Como os usuários em potencial descobrirão seus dados?</li> <li>– Com quem você compartilhará os dados e em que condições?</li> <li>– Você compartilhará dados por meio de um repositório, lidará com solicitações diretamente ou usará outro mecanismo?</li> <li>– Quando você disponibilizará os dados?</li> <li>– Você pretende obter um identificador persistente para seus dados?</li> </ul>	Considere onde, como e para quem os dados com valor reconhecido a longo prazo devem ser disponibilizados. Os métodos usados para compartilhar dados dependerão de vários fatores, como tipo, tamanho, complexidade e sensibilidade dos dados. Se possível, mencione exemplos anteriores para mostrar um histórico de compartilhamento eficaz de dados. Considere como as pessoas podem reconhecer a reutilização de seus dados.
<b>Alguma restrição no compartilhamento de dados é necessária?</b>	<ul style="list-style-type: none"> <li>– Que ação você tomará para superar ou minimizar as restrições?</li> <li>– Por quanto tempo você precisa de uso exclusivo dos dados e por quê?</li> <li>– Será necessário um contrato de compartilhamento de dados (ou equivalente)?</li> </ul>	Descreva quaisquer dificuldades esperadas no compartilhamento de dados com valor reconhecido a longo prazo, juntamente com as causas e possíveis medidas para superá-las. As restrições podem ser devido à confidencialidade, à falta de acordos de consentimento ou ao DPI, por exemplo. Considere se um acordo de confidencialidade daria proteção suficiente para dados confidenciais.
<b>Responsabilidades e recursos</b>		
<b>Quem será responsável pelo gerenciamento de dados?</b>	<ul style="list-style-type: none"> <li>– Quem é responsável por implementar o DMP e garantir que ele seja revisado e revisado?</li> <li>– Quem será responsável por cada atividade de gerenciamento de dados?</li> <li>– Como as responsabilidades serão divididas entre os sites parceiros em projetos de pesquisa colaborativa?</li> </ul>	Descreva as funções e as responsabilidades para todas as atividades, por exemplo, captura de dados, produção de metadados, qualidade de dados, armazenamento e <i>backup</i> , arquivamento e compartilhamento de dados. Considere quem será responsável por garantir que as políticas relevantes sejam respeitadas. Os indivíduos devem ser nomeados sempre que possível.

	<ul style="list-style-type: none"> <li>– A propriedade dos dados e as responsabilidades pelo RDM farão parte de algum acordo de consórcio ou contrato acordado entre os parceiros?</li> </ul>	
<p><b>Quais recursos você precisará para entregar seu plano?</b></p>	<ul style="list-style-type: none"> <li>– É necessária especialização adicional (ou treinamento para o pessoal existente)?</li> <li>– Você precisa de <i>hardware</i> ou <i>software</i> que seja adicional ou excepcional à provisão institucional existente?</li> <li>– As cobranças serão aplicadas pelos repositórios de dados?</li> </ul>	<p>Considere cuidadosamente quaisquer recursos necessários para entregar o plano, por exemplo, <i>software</i>, <i>hardware</i>, conhecimento técnico etc. Quando recursos dedicados são necessários, eles devem ser descritos e justificados.</p>

## APÊNDICE D – POLÍTICA DE GESTÃO DE DADOS DE PESQUISA

Institui ações para a gestão e preservação de dados e o repositório de dados de pesquisa na universidade.

### CAPÍTULO I – DEFINIÇÃO

Art. 1 – Os dados de pesquisa são todos os insumos coletados/produzidos dentro dos laboratórios de pesquisa e/ou produzidos nas pesquisas de teses ou dissertações, independente do formato ou do conteúdo, e devem ser disponibilizados na íntegra no repositório de dados.

I – **Dados de pesquisa incluem:** documentos textuais, planilhas, estatísticas, cadernos de laboratório, cadernos de campo, diários, questionários, transcrições, arquivos de áudio, vídeo, fotografias, sequências de proteínas ou genéticos, artefatos, amostras, modelos, algoritmos, *scripts*, arquivos de *log*, *software* de simulação, metodologias, fluxos de trabalho, procedimentos operacionais, padrões e protocolos.

§ 1 – Os dados coletados variam de acordo com a área de conhecimento.

Art. 2 – Institui a obrigatoriedade da gestão de dados de pesquisa na instituição.

Parágrafo único – A gestão de dados compreende ações necessárias para a qualidade dos dados de pesquisa. É obrigatório adotar um modelo de ciclo de vida de dados pelos gestores de pesquisas, realizar a descrição do Plano de Gestão de Dados, descrever o glossário de termos usados, realizar a preservação dos dados na fase intermediária e o depósito no repositório de dados da instituição.

Art. 3 – Institui o repositório de dados de pesquisa na universidade.

I – O repositório de dados é um serviço mantido pela instituição para o depósito dos dados produzidos pelos seus pesquisadores (professores, técnicos e alunos: graduação, mestrado e doutorado).

§1 – O repositório de dados da instituição é uma plataforma de *software* livre que cumpre o protocolo (OAIS).

Art. 4 – Institui o *software* de preservação digital para o repositório de dados de pesquisa.

§1 – O *software* de preservação digital deve ser uma plataforma de *software* livre que cumpre o protocolo (OAIS).

§2 – Deverá ser possível a integração com a plataforma de *software* de repositório escolhido.

§3 – O repositório deve adotar identificadores persistentes para serem atribuídos aos conjuntos de dados.

## **CAPÍTULO II – DOS BENEFÍCIOS**

Art. 5 – O compartilhamento de dados é uma exigência científica ao nível global, conduzindo a ciência a novas descobertas.

Art. 6 – Os conjuntos de dados são insumos valiosos e seu compartilhamento se constitui como produção científica, assim, o pesquisador também recebe citação dos seus dados.

Art. 7 – Permite que suas pesquisas sejam refeitas, testadas e validadas por outros pesquisadores, dando mais garantia de autenticidade, qualidade, integridade e ética científica.

## **CAPÍTULO III – DAS RESPONSABILIDADES**

Art. 8 – Da biblioteca universitária:

I – Presidir o Comitê de Gestão de Dados.

II – Desenvolver uma estrutura de serviços para apoio à gestão de dados.

III – Construir e gerenciar um *site* de orientação para a comunidade universitária sobre boas práticas de gestão de dados

*a) Construir uma página de perguntas frequentes sobre a gestão e curadoria de dados;*

IV - Estruturar o setor/departamento de Serviço de Suporte a Dados de pesquisa.

*a) Oferecer atendimento de referência sobre todos os temas relacionados à gestão de dados;*

*b) Realizar treinamentos sobre a gestão de dados: ciclo de vida dos dados, Plano de Gestão de Dados (PGD), ferramentas para elaboração do PGD, tipos de licenças para o conjunto de dados, como citar e reusar dados.*

*c) Realizar a gestão e a curadoria do repositório de dados da instituição.*

Art. 9 – Dos pesquisadores:

§1 – A responsabilidade em adotar e seguir um modelo de ciclo de vida dos dados de pesquisa.

§2 – São responsáveis por descreverem os Planos de Gestão de Dados (PGD) e atualizá-los ao longo do projeto.

*a) Indicar todas as publicações e onde estão localizados os dados de pesquisa.*

*b) Utilizar a ferramenta DMPTool ou DMPonline.*

§3 – Produzir um documento em separado chamado “Glossário de Termos” explicando as definições e as variáveis do conjunto de dados.

*a) Este deve ser armazenado junto ao conjunto de dados e ao PGD.*

§4 – Ao registrarem/detalhar os metadados descritivos do conjunto de dados, utilizar padrões internacionais de metadados, tornando-os disponíveis e detectáveis para outros pesquisadores de uma forma que os ajude a entender a pesquisa e o potencial de reutilização dos dados.

§5 – Indicar os resultados publicados que utilizam aqueles conjuntos de dados depositados no repositório.

Art. 10 – Da Pró-Reitoria de Pesquisa:

I – Ser membro do Comitê de Gestão de Dados.

II – Responsável por identificar os laboratórios de pesquisa existentes na instituição.

III – Responsável por conhecer todos os tipos, formatos e volumes de dados produzidos nos laboratórios da instituição.

IV – Instituir a obrigatoriedade na elaboração do Plano de Gestão de Dados (PGD) dos professores e pesquisadores da instituição.

V – Incentivar que os professores realizem a gestão e o armazenamento de dados no repositório da instituição.

Art. 11 – Da Pró-Reitoria de Pós-Graduação:

I – Ser membro do Comitê de Gestão de Dados.

II – Incentivar a gestão de dados de pesquisas nos trabalhos de mestrado e doutorado.

III – Instituir a obrigatoriedade na elaboração do Plano de Gestão de Dados (PGD) dos estudantes de mestrado e doutorado para sua comunidade universitária.

IV – Incentivar que os alunos da pós-graduação realizem os treinamentos oferecidos pelo Serviço de Suporte à Pesquisa, ministrados pela biblioteca.

Art. 12 – Do Departamento de Tecnologia da Informação:

I – Ser membro do Comitê de Gestão de Dados.

- II – Hospedar a infraestrutura tecnológica do repositório de dados.
- III – Instalar e customizar a plataforma de repositório.
- IV – Implementar *software* de preservação digital com modelo (OAIS).
- V – Realizar as atualizações dos *softwares* do repositório e o de preservação, quando necessário.
- VI – Manter acesso contínuo ao repositório de dados da instituição.

Art. 13 – Da Universidade/Reitoria:

- I – Aprovar e publicar atos normativos/portarias instituindo a gestão de dados de pesquisa na instituição.
- II - Responsável por fornecer orçamento específico para o repositório de dados.
  - a) *Os recursos devem ser utilizados para a capacitação da equipe, recursos de TI, recursos para identificadores persistentes e demais custos.*

#### **CAPÍTULO IV – DO DEPÓSITO DOS DADOS DE PESQUISA**

Art. 14 – Os dados de pesquisa devem ser depositados no repositório de dados da instituição.

§1 - Os dados sensíveis devem ser desidentificados, deve-se proteger a privacidade dos sujeitos e indivíduos.

- a) *Constituem-se como dados sensíveis aqueles que podem identificar/localizar pessoas, como: nome, endereço, profissão e local de trabalho. É todo tipo de registro que consegue ao ser analisado com outras variáveis determinar quem é aquele sujeito.*

§2 – Os dados embargados devem cumprir o prazo de embargo, porém, seus metadados de identificação devem estar disponíveis para consulta no repositório.

- a) *É necessário declarações que expliquem o motivo pelas restrições, com motivos relevantes na publicação da pesquisa e na descrição dos metadados do conjunto de dados embargado.*

§3 – Os dados devem estar de acordo com os princípios de dados FAIR: localizável, acessível, interoperável e reutilizável.

§4 – Dados financiados com dinheiro público de agências de fomento e/ou provenientes de universidades públicas, mesmo que a pesquisa não tenha sido financiada diretamente, devem ser depositados em acesso aberto.

- a) *Os dados de pesquisa financiados são um bem público e produzidos no interesse público.*

Art. 15 – Os dados de pesquisa depositados em repositórios internacionais por exigência de periódicos internacionais também devem ser depositados no repositório da instituição

I – Em virtude de regra do periódico internacional restringindo esse novo depósito, deve-se constar os metadados no repositório da instituição, possibilitando o acesso ao conjunto de dados. O repositório internacional deve estar em acesso aberto.

## **CAPÍTULO V – DA PROPRIEDADE INTELECTUAL**

Art. 16 – A propriedade intelectual dos dados gerados na instituição é dos seus pesquisadores ou da instituição. Os direitos autorais são irrenunciáveis e inalienáveis.

Art. 17 – É obrigatória a atribuição de licenças abertas para o conjunto ou arquivo de dados depositados no repositório de dados.

I – É recomendado utilizar as licenças *Creative Commons* ou *Open Data Commons*.

## **CAPÍTULO VI – DO ARMAZENAMENTO DE DADOS**

Art. 18 – Durante o curso da pesquisa, o pesquisador é o responsável pelo armazenamento seguro dos dados em ambiente seguro, recomenda:

- a) *O backup dos dados (diário, semanal ou mensal), levando em consideração o volume coletado para decidir.*
- b) *Utilizar do armazenamento em nuvem.*

Art. 19 – Após o encerramento da pesquisa, os arquivos de dados ou os conjuntos de dados deverão ser depositados em formato aberto em um repositório de dados.

I – É obrigatório depositar o Plano de Gestão de Dados (PGD) para cada conjunto de dados.

II – É obrigatório depositar um “Glossário de Termos” explicando as variáveis e as expressões utilizadas para o entendimento do conjunto de dados.

III – É obrigatório o preenchimento dos metadados descritivos do repositório de dados.

IV – Os dados devem ser depositados em até 30 dias após a conclusão da pesquisa.

- a) *Cada conjunto de dados deve receber um identificador persistente.*

Art. 20 – Os dados devem ser armazenados por **pelo menos 10 anos** ou por **10 anos a partir da última solicitação** de acesso.



## **CAPÍTULO VII – DOS METADADOS**

Art. 21 – Os metadados são importantes para fins de descrição, citação, descoberta e recuperação de coleções de dados de pesquisa.

Art. 22 – Os metadados deverão seguir padrões internacionais.

I – É de responsabilidade do pesquisador ou do líder do repositório realizar a descrição dos metadados.

§1 – Os metadados descritivos são aqueles que devem ser preenchidos para que os conjuntos de dados sejam reusáveis.

§2 – Os metadados administrativos são aqueles que explicam o ciclo de vida dos dados, a forma da coleta, a data, os instrumentos e a metodologia.

§3 – Os metadados estruturais são aqueles que explicam a estrutura do conjunto de dados.

## **CAPÍTULO VIII – DO REUSO DOS DADOS**

Art. 23 – O reuso dos dados se dá quando o pesquisador utiliza dados coletados por outros pesquisadores.

I – É obrigatório realizar a citação do conjunto de dados indicando onde está depositado e quem são os autores dos dados.

Art. 24 – Os produtos provenientes do reuso de dados de pesquisa, como os artigos, as teses, as dissertações e/ou novos conjuntos de dados devem ser compartilhados com a mesma licença e as mesmas condições.

## **CAPÍTULO IX – DAS PENALIDADES**

Art. 25 – O pesquisador não receberá financiamento caso se recuse a compartilhar seus dados de pesquisa.

I – O arquivo ou o conjunto de dados devem estar em formato aberto para o reuso.

a) *Deverá constar toda a documentação como: Plano de Gestão de dados (PGD) e Glossário de Termos.*

II – Fica proibido compartilhar arquivo ou conjunto de dados em formato PDF.

Art. 26 – O pesquisador que se recusar ou dificultar a disponibilizar seus dados ou de orientandos não receberá financiamentos e bolsas do programa de pós-graduação que estiver vinculado.

## ANEXO A – PADRÕES DE METADADOS

<b>Área do conhecimento</b>	<b>Padrão de metadados</b>	<b>Informações</b>
Geral	CERIF-Common European Research Information Format	<a href="http://www.dcc.ac.uk/resources/metadata-standards/cerif-common-european-research-information-format">http://www.dcc.ac.uk/resources/metadata-standards/cerif-common-european-research-information-format</a>
Geral	Data Package	<a href="http://www.dcc.ac.uk/resources/metadata-standards/data-package">http://www.dcc.ac.uk/resources/metadata-standards/data-package</a>
Geral	DataCite Metadata Schema	<a href="http://www.dcc.ac.uk/resources/metadata-standards/datacite-metadata-schema">http://www.dcc.ac.uk/resources/metadata-standards/datacite-metadata-schema</a>
Geral	DCAT-Data Catalog Vocabulary	<a href="http://www.dcc.ac.uk/resources/metadata-standards/dcat-data-catalog-vocabulary">http://www.dcc.ac.uk/resources/metadata-standards/dcat-data-catalog-vocabulary</a>
Geral	Dublin Core	<a href="http://www.dcc.ac.uk/resources/metadata-standards/dublin-core">http://www.dcc.ac.uk/resources/metadata-standards/dublin-core</a>
Geral	OAI-ORE-Open Archives Initiative Object Reuse and Exchange	<a href="http://www.dcc.ac.uk/resources/metadata-standards/oai-ore-open-archives-initiative-object-reuse-and-exchange">http://www.dcc.ac.uk/resources/metadata-standards/oai-ore-open-archives-initiative-object-reuse-and-exchange</a>
Geral	Observations and Measurements	<a href="http://www.dcc.ac.uk/resources/metadata-standards/observations-and-measurements">http://www.dcc.ac.uk/resources/metadata-standards/observations-and-measurements</a>
Geral	PREMIS	<a href="http://www.dcc.ac.uk/resources/metadata-standards/premis">http://www.dcc.ac.uk/resources/metadata-standards/premis</a>
Geral	PROV	<a href="http://www.dcc.ac.uk/resources/metadata-standards/prov">http://www.dcc.ac.uk/resources/metadata-standards/prov</a>
Geral	RDF Data Cube Vocabulary	<a href="http://www.dcc.ac.uk/resources/metadata-standards/rdf-data-cube-vocabulary">http://www.dcc.ac.uk/resources/metadata-standards/rdf-data-cube-vocabulary</a>
Geral	Repository-Developed Metadata Schemas	<a href="http://www.dcc.ac.uk/resources/metadata-standards/repository-developed-metadata-schemas">http://www.dcc.ac.uk/resources/metadata-standards/repository-developed-metadata-schemas</a>
Biologia	ABCD-Access to Biological Collection Data	<a href="http://www.dcc.ac.uk/resources/metadata-standards/abcd-access-biological-collection-data">http://www.dcc.ac.uk/resources/metadata-standards/abcd-access-biological-collection-data</a>
Biologia	Darwin Core	<a href="http://www.dcc.ac.uk/resources/metadata-standards/darwin-core">http://www.dcc.ac.uk/resources/metadata-standards/darwin-core</a>
Biologia	EML-Ecological Metadata Language	<a href="http://www.dcc.ac.uk/resources/metadata-standards/eml-ecological-metadata-language">http://www.dcc.ac.uk/resources/metadata-standards/eml-ecological-metadata-language</a>
Biologia	Genome Metadata	<a href="http://www.dcc.ac.uk/resources/metadata-standards/genome-metadata">http://www.dcc.ac.uk/resources/metadata-standards/genome-metadata</a>
Biologia	ISA-Tab	<a href="http://www.dcc.ac.uk/resources/metadata-standards/isa-tab">http://www.dcc.ac.uk/resources/metadata-standards/isa-tab</a>
Biologia	MIBBI-Minimum Information for Biological and Biomedical Investigations	<a href="http://www.dcc.ac.uk/resources/metadata-standards/mibbi-minimum-information-biological-and-biomedical-investigations">http://www.dcc.ac.uk/resources/metadata-standards/mibbi-minimum-information-biological-and-biomedical-investigations</a>
Biologia	Observ-OM	<a href="http://www.dcc.ac.uk/resources/metadata-standards/observ-om">http://www.dcc.ac.uk/resources/metadata-standards/observ-om</a>
Biologia	OME-XML-Open Microscopy Environment XML	<a href="http://www.dcc.ac.uk/resources/metadata-standards/ome-xml-open-microscopy-environment-xml">http://www.dcc.ac.uk/resources/metadata-standards/ome-xml-open-microscopy-environment-xml</a>
Biologia	PDBx/mmCIF-Protein Data Bank Exchange Dictionary and the Macromolecular Crystallographic Information Framework	<a href="http://www.dcc.ac.uk/resources/metadata-standards/pdbxmmcif-%E2%80%93-protein-data-bank-exchange-dictionary-and-macromolecular-cr">http://www.dcc.ac.uk/resources/metadata-standards/pdbxmmcif-%E2%80%93-protein-data-bank-exchange-dictionary-and-macromolecular-cr</a>
Biologia	Protocol Data Element Definitions	<a href="http://www.dcc.ac.uk/resources/metadata-standards/protocol-data-element-definitions">http://www.dcc.ac.uk/resources/metadata-standards/protocol-data-element-definitions</a>
Biologia	Repository-Developed Metadata Schemas	<a href="http://www.dcc.ac.uk/resources/metadata-standards/repository-developed-metadata-schemas">http://www.dcc.ac.uk/resources/metadata-standards/repository-developed-metadata-schemas</a>
Ciências da Terra	AgMES-Agricultural Metadata Element Set	<a href="http://www.dcc.ac.uk/resources/metadata-standards/agmes-agricultural-metadata-element-set">http://www.dcc.ac.uk/resources/metadata-standards/agmes-agricultural-metadata-element-set</a>
Ciências da Terra	AVM-Astronomy Visualization Metadata	<a href="http://www.dcc.ac.uk/resources/metadata-standards/avm-astronomy-visualization-metadata">http://www.dcc.ac.uk/resources/metadata-standards/avm-astronomy-visualization-metadata</a>

Ciências da Terra	CF (Climate and Forecast) Metadata Conventions	<a href="http://www.dcc.ac.uk/resources/metadata-standards/cf-climate-and-forecast-metadata-conventions">http://www.dcc.ac.uk/resources/metadata-standards/cf-climate-and-forecast-metadata-conventions</a>
Ciências da Terra	CIM - Common Information Model	<a href="http://www.dcc.ac.uk/resources/metadata-standards/cim-common-information-model">http://www.dcc.ac.uk/resources/metadata-standards/cim-common-information-model</a>
Ciências da Terra	DIF -Directory Interchange Format	<a href="http://www.dcc.ac.uk/resources/metadata-standards/dif-directory-interchange-format">http://www.dcc.ac.uk/resources/metadata-standards/dif-directory-interchange-format</a>
Ciências da Terra	FGDC/CSDGM-Federal Geographic Data Committee Content Standard for Digital Geospatial Metadata	<a href="http://www.dcc.ac.uk/resources/metadata-standards/fgdcssdgm-federal-geographic-data-committee-content-standard-digital-ge">http://www.dcc.ac.uk/resources/metadata-standards/fgdcssdgm-federal-geographic-data-committee-content-standard-digital-ge</a>
Ciências da Terra	ISO 19115	<a href="http://www.dcc.ac.uk/resources/metadata-standards/iso-19115">http://www.dcc.ac.uk/resources/metadata-standards/iso-19115</a>
Ciências da Terra	Observations and Measurements	<a href="http://www.dcc.ac.uk/resources/metadata-standards/observations-and-measurements">http://www.dcc.ac.uk/resources/metadata-standards/observations-and-measurements</a>
Ciências da Terra	Repository-Developed Metadata Schemas	<a href="http://www.dcc.ac.uk/resources/metadata-standards/repository-developed-metadata-schemas">http://www.dcc.ac.uk/resources/metadata-standards/repository-developed-metadata-schemas</a>
Ciências Físicas	AVM-Astronomy Visualization Metadata	<a href="http://www.dcc.ac.uk/resources/metadata-standards/avm-astronomy-visualization-metadata">http://www.dcc.ac.uk/resources/metadata-standards/avm-astronomy-visualization-metadata</a>
Ciências Físicas	CIF -Crystallographic Information Framework	<a href="http://www.dcc.ac.uk/resources/metadata-standards/cif-crystallographic-information-framework">http://www.dcc.ac.uk/resources/metadata-standards/cif-crystallographic-information-framework</a>
Ciências Físicas	CSMD-CCLRC Core Scientific Metadata Model	<a href="http://www.dcc.ac.uk/resources/metadata-standards/csmd-cclrc-core-scientific-metadata-model">http://www.dcc.ac.uk/resources/metadata-standards/csmd-cclrc-core-scientific-metadata-model</a>
Ciências Físicas	FITS -Flexible Image Transport System	<a href="http://www.dcc.ac.uk/resources/metadata-standards/fits-flexible-image-transport-system">http://www.dcc.ac.uk/resources/metadata-standards/fits-flexible-image-transport-system</a>
Ciências Físicas	International Virtual Observatory Alliance Technical Specifications	<a href="http://www.dcc.ac.uk/resources/metadata-standards/international-virtual-observatory-alliance-technical-specifications">http://www.dcc.ac.uk/resources/metadata-standards/international-virtual-observatory-alliance-technical-specifications</a>
Ciências Físicas	NeXus	<a href="http://www.dcc.ac.uk/resources/metadata-standards/nexus">http://www.dcc.ac.uk/resources/metadata-standards/nexus</a>
Ciências Físicas	Observations and Measurements	<a href="http://www.dcc.ac.uk/resources/metadata-standards/observations-and-measurements">http://www.dcc.ac.uk/resources/metadata-standards/observations-and-measurements</a>
Ciências Físicas	PDBx/mmCIF –Protein Data Bank Exchange Dictionary and the Macromolecular Crystallographic Information Framework	<a href="http://www.dcc.ac.uk/resources/metadata-standards/pdbxmmcif-%E2%80%93protein-data-bank-exchange-dictionary-and-macromolecular-cr">http://www.dcc.ac.uk/resources/metadata-standards/pdbxmmcif-%E2%80%93protein-data-bank-exchange-dictionary-and-macromolecular-cr</a>
Ciências Físicas	SDAC-Standard for Documentation of Astronomical Catalogues	<a href="http://www.dcc.ac.uk/resources/metadata-standards/sdac-standard-documentation-astronomical-catalogues">http://www.dcc.ac.uk/resources/metadata-standards/sdac-standard-documentation-astronomical-catalogues</a>
Ciências Físicas	SPASE Data Model	<a href="http://www.dcc.ac.uk/resources/metadata-standards/spase-data-model">http://www.dcc.ac.uk/resources/metadata-standards/spase-data-model</a>
Ciências Sociais e Humanidades	DDI-Data Documentation Initiative	<a href="http://www.dcc.ac.uk/resources/metadata-standards/ddi-data-documentation-initiative">http://www.dcc.ac.uk/resources/metadata-standards/ddi-data-documentation-initiative</a>
Ciências Sociais e Humanidades	MIDAS-Heritage	<a href="http://www.dcc.ac.uk/resources/metadata-standards/midas-heritage">http://www.dcc.ac.uk/resources/metadata-standards/midas-heritage</a>
Ciências Sociais e Humanidades	OAI-ORE-Open Archives Initiative Object Reuse and Exchange	<a href="http://www.dcc.ac.uk/resources/metadata-standards/oai-ore-open-archives-initiative-object-reuse-and-exchange">http://www.dcc.ac.uk/resources/metadata-standards/oai-ore-open-archives-initiative-object-reuse-and-exchange</a>
Ciências Sociais e Humanidades	QuDEx-Qualitative Data Exchange Format	<a href="http://www.dcc.ac.uk/resources/metadata-standards/qudex-qualitative-data-exchange-format">http://www.dcc.ac.uk/resources/metadata-standards/qudex-qualitative-data-exchange-format</a>
Ciências Sociais e Humanidades	SDMX-Statistical Data and Metadata Exchange	<a href="http://www.dcc.ac.uk/resources/metadata-standards/sdmx-statistical-data-and-metadata-exchange">http://www.dcc.ac.uk/resources/metadata-standards/sdmx-statistical-data-and-metadata-exchange</a>

Fonte: Elaborado por Almeida (2019) com base no DCC.