

UNIVERSIDADE FEDERAL DE MINAS GERAIS
ESCOLA DE CIÊNCIA DA INFORMAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM GESTÃO & ORGANIZAÇÃO DO
CONHECIMENTO

Fernanda Gomes Almeida

SUPORTE À GESTÃO DE DADOS DE PESQUISA: uma ampliação dos serviços
oferecidos pelas bibliotecas

Belo Horizonte
2019

Fernanda Gomes Almeida

**SUPORTE À GESTÃO DE DADOS DE PESQUISA: uma ampliação dos serviços
oferecidos pelas bibliotecas**

Versão final

Tese apresentada ao Programa de Pós-Graduação em Gestão & Organização do Conhecimento da Universidade Federal de Minas Gerais, como requisito parcial à obtenção do grau de Doutor, área de concentração Ciência da Informação.

Linha de Pesquisa: Gestão e Tecnologia

Orientadora: Prof.^a Dr^a Beatriz Valadares Cendón

Belo Horizonte

2019

Almeida, Fernanda Gomes.

A447s Suporte à gestão de dados de pesquisa [recurso eletrônico] : uma ampliação dos serviços oferecidos pelas bibliotecas / Fernanda Gomes Almeida. ĩ 2019.
1 recurso online (249 f. : il., color.) : pdf.

Orientadora: Beatriz Valadares Cendón

Tese (doutorado) ĩ Universidade Federal de Minas Gerais, Escola de Ciência da Informação.

Referências: f. 197-218.

Apêndices: f. 219-249.

Exigências do sistema: Adobe Acrobat Reader.

1. Ciência da informação ĩ Teses. 2. Processamento eletrônico de dados na pesquisa ĩ Teses 3. Bibliotecas universitárias ĩ Serviços de estatística ĩ Teses. I. Título. II. Cendón, Beatriz Valadares. III. Universidade Federal de Minas Gerais, Escola de Ciência da Informação.

CDU: 025.5:004.65



FOLHA DE APROVAÇÃO

Suporte à gestão de dados de pesquisa: uma ampliação dos serviços oferecidos pelas bibliotecas

FERNANDA GOMES ALMEIDA

Tese submetida à Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em GESTÃO E ORGANIZAÇÃO DO CONHECIMENTO, como requisito para obtenção do grau de Doutor em GESTÃO E ORGANIZAÇÃO DO CONHECIMENTO, área de concentração CIÊNCIA DA INFORMAÇÃO, linha de pesquisa Gestão e Tecnologia.

Aprovada em 06 de agosto de 2019, pela banca constituída pelos membros:

Prof(a). Beatriz Valadares Cendón (Orientadora)
Aposentada/UFGM

Prof(a). Luana Farias Sales Marques
IBICT [por videoconferência]

Prof(a). Ricardo Rodrigues Barbosa
Aposentado/UFGM

Prof(a). Wellington Marçal de Carvalho
UNIR

Prof(a). Consuelo Chaves Joncew
Aposentada/CEF

Dr(a). Wesley Rodrigo Fernandes
FALE/UFGM

Belo Horizonte, 6 de agosto de 2019.



ATA DA DEFESA DE TESE DA ALUNA FERNANDA GOMES ALMEIDA

Realizou-se, no dia 06 de agosto de 2019, às 14:00 horas, Biblioteca Central da UFMG - Auditório do 4º andar, da Universidade Federal de Minas Gerais, a defesa de tese, intitulada *Suporte à gestão de dados de pesquisa: uma ampliação dos serviços oferecidos pelas bibliotecas*, apresentada por FERNANDA GOMES ALMEIDA, número de registro 2015665638, graduada no curso de BIBLIOTECONOMIA, como requisito parcial para a obtenção do grau de Doutor em GESTÃO E ORGANIZAÇÃO DO CONHECIMENTO, à seguinte Comissão Examinadora: Prof(a). Beatriz Valadares Cendón - Aposentada/UFMG (Orientadora), Prof(a). Luana Farias Sales Marques - IBICT [por videoconferência], Prof(a). Ricardo Rodrigues Barbosa - Aposentado/UFMG, Prof(a). Wellington Marçal de Carvalho - UNIR, Prof(a). Consuelo Chaves Joncew - Aposentada/CEF, Dr(a). Wesley Rodrigo Fernandes - FALE/UFMG. A banca recomendou a publicação da tese para os prêmios da área e também para o prêmio capes.

A Comissão considerou a tese:

☒ Aprovada

☐ Reprovada

Finalizados os trabalhos, lavrei a presente ata que, lida e aprovada, vai assinada por mim e pelos membros da Comissão.

Belo Horizonte, 06 de agosto de 2019.

Prof(a). Beatriz Valadares Cendón

Prof(a). Luana Farias Sales Marques

Prof(a). Ricardo Rodrigues Barbosa

Prof(a). Wellington Marçal de Carvalho

Prof(a). Consuelo Chaves Joncew

Dr(a). Wesley Rodrigo Fernandes

*Aos meus pais, João Valadão de Almeida e Maria Gomes Faria de Almeida; ao meu irmão,
Fabiano Faria Valadão.
Com amor!*

AGRADECIMENTOS

A Deus, pela vida, por me dar força e coragem! À Universidade Federal de Minas Gerais, pela oportunidade de formação e crescimento na graduação, mestrado e doutorado e por me acolher como servidora técnico-administrativa.

Aos meus pais, Seu João Valadão e Dona Neném. Antes de tudo, sou a filha deles! Obrigada por todo amor, carinho, dedicação, privações e renúncia! Esta é mais uma vitória de vocês e para vocês! Ao meu irmão, Fabiano: amigo, parceiro e incentivador! À minha família: Valadão de Almeida e Gomes de Faria!

À minha orientadora, Professora Beatriz Valadares Cendón, pela oportunidade, apoio, atenção e ensinamentos durante a longa trajetória de mestrado e doutorado.

Aos professores que participaram da banca de qualificação e defesa: Luís Fernando Sayão, Maurício Barcellos de Almeida, Ricardo Rodrigues Barbosa, Luana Farias Sales, Wellington Marçal de Carvalho, Consuelo Chaves Joncew, Wesley Rodrigo Fernandes. Obrigada pelas críticas e valiosas contribuições que enriqueceram este trabalho.

Aos professores, servidores e colegas do Programa de Pós-graduação em Ciência da Informação e do Programa de Pós-graduação em Gestão & Organização do Conhecimento da UFMG.

À Diretoria da Biblioteca Universitária da UFMG - Wellington Marçal de Carvalho e Anália das Graças Gandini Pontelo, seguida por Kátia Lúcia Pacheco e Sindier Antônia Alves -, pela concessão de afastamento.

À família Azevedo Costa que me recebeu com tanto carinho: Maurina, e as amigas-irmãs doutorandas, Aline e Mirna.

Aos amigos da Biblioteca Universitária, Biblioteca Central da UFMG e seus “inquilinos”.

À Leila Anastácio e Maria Clara Caldas, amigas doutorandas e colegas de casa.

Ao Marcos de Souza, pelo incentivo e por me ensinar a levar uma vida mais leve!

Aos amigos que a corrida me deu, @clubeendorfina.

A todos aqueles que contribuíram para a realização deste trabalho, que estiveram comigo, incentivaram, sofreram e se alegraram com esta vitória!

A todos, obrigada!!!

RESUMO

Pesquisa qualitativa que teve como objetivo a proposição de uma taxonomia de serviços voltados à gestão dos dados de pesquisa gerados nas universidades. Além disso, a pesquisa buscou sintetizar dificuldades, desafios e ações necessárias para viabilização e manutenção desses serviços nas instituições. O trabalho traz, como fundamentação teórica, os dados de pesquisa, o paradigma da *eScience*, *big science* e *small science* e as diferenças na geração e estruturação de dados, gestão de dados de pesquisa, ciclo de vida dos dados, compartilhamento e reuso de dados, publicação e citação, infraestruturas, padrões e políticas para dados, as bibliotecas e a gestão de dados de pesquisa, incluindo os desafios enfrentados para a entrega de serviços de apoio à gestão de dados pelas bibliotecas, as práticas e necessidades dos pesquisadores e, finalmente, os Serviços de Suporte à Gestão de Dados de Pesquisa (SSGDPs). O trabalho constou de três etapas: 1) identificação de serviços, ações e dificuldades e desafios por meio da literatura; 2) identificação e contextualização de serviços e identificação das ações e dificuldades por meio de entrevistas; e 3) análise dos dados. A coleta de dados aconteceu nas etapas 1 e 2. Na etapa 1 optou-se por pesquisa bibliográfica em materiais já publicados, identificados a partir de buscas em bases de dados, Google Acadêmico e referências citadas. Foram selecionados 165 artigos. Na etapa 2, a coleta de dados ocorreu por meio de entrevistas. Foram entrevistadas duas instituições selecionadas a partir de uma amostra não probabilística intencional. A identificação das instituições foi realizada por meio de pesquisa na *web*, e após isso, pelo envio de carta-convite ou pedido de informações a respeito da disponibilização ou participação das bibliotecas em discussões para o desenvolvimento de serviços ou infraestruturas. Assim, participaram das entrevistas: a Universidade Federal de Santa Catarina (UFSC) e a Universidade Federal do Rio Grande do Sul (UFRGS). A análise de conteúdo proposta por Bardin (1977) foi usada como método para o tratamento e interpretação dos dados coletados nas etapas 1 (literatura) e 2 (entrevistas). Os 165 artigos e as duas entrevistas foram lidos integralmente. Todo o material foi codificado e categorizado com o auxílio do *software* NVivo. Conjuntamente, os dados coletados nas etapas 1 e 2 (literatura e entrevistas) permitiram: 1) a sistematização e proposição de uma taxonomia de serviços estrutura em três categorias principais (Infraestrutura, Suporte e Educacional) que são subdivididas nas subcategorias 'Gestão e curadoria de dados' e 'Outros'. Essas duas subcategorias têm subdivisões adicionais que são replicadas para cada uma das categorias principais. Os serviços categorizados como Suporte foram os mais frequentes nos 167 documentos (165 artigos e duas entrevistas) utilizados para a construção da taxonomia. Em seguida aparecem os serviços da categoria Educacional e, por último, os serviços da categoria Infraestrutura. As duas categorias mais presentes – Suporte e Educacional – compreendem principalmente os serviços de referência, instrução e orientação, representando extensão dos serviços tradicionais já realizados pelas bibliotecas, agora com o foco nos dados de pesquisa; 2) a identificação e compilação dos serviços de suporte à gestão de dados de pesquisa oferecidos; 3) a exposição dos desafios e dificuldades enfrentadas nas instituições entrevistadas; e 4) a síntese das ações gerais para a viabilização e manutenção dos serviços nas instituições. A etapa 2 (coleta de dados por entrevistas) permitiu, ainda, a: 5) contextualização dos esforços para o desenvolvimento de SSGDPs nas instituições entrevistadas. Consideramos que a pesquisa cumpriu os objetivos propostos de identificar serviços, ações, dificuldades e desafios; contextualizar o oferecimento de serviços e infraestruturas nas universidades entrevistadas; sistematizar e propor uma taxonomia de serviços; sintetizar as dificuldades e desafios; e sintetizar as ações para viabilização e manutenção dos serviços nas instituições.

Palavras-chave: Dados de pesquisa; Gestão de dados de pesquisa; Serviços de suporte à gestão de dados; Bibliotecas; Serviços.

ABSTRACT

Qualitative research aimed at proposing a taxonomy of services focused on the management of research data generated in universities. In addition, the research sought to synthesize difficulties, challenges and actions necessary for the viability and maintenance of these services in the institutions. The work brings, as a theoretical foundation, the research data, the eScience paradigm, big science and small science and the differences in data generation and structuring, research data management, data life cycle, data sharing and reuse, publication and citation, infrastructures, standards and policies for data, libraries and research data management, including the challenges faced for the delivery of data management support services by libraries, the practices and needs of researchers and, finally, the Research Data Management Support Services (RDMSSs). The work consisted of three steps: 1) identification of services, actions and challenges through literature; 2) identification and contextualization of services and identification of actions and challenges through interviews; and 3) data analysis. Data collection took place in stages 1 and 2. In step 1 we opted for bibliographic search in materials already published, identified from searches in databases, Google Academic and cited references. A total of 165 articles were selected. In step 2, data collection occurred through interviews. Two institutions selected from an intentional non-probability sample were interviewed. The identification of the institutions was carried out by means of web research, and after that, by sending a letter of invitation or request for information regarding the availability or participation of libraries in discussions for the development of services or infrastructure. Thus, the Federal University of Santa Catarina (UFSC) and the Federal University of Rio Grande do Sul (UFRGS) participated in the interviews. The content analysis proposed by Bardin (1977) was used as a method for the treatment and interpretation of the data collected in steps 1 (literature) and 2 (interviews). The 165 articles and the two interviews were read in full. All the material was coded and categorized with the help of NVivo software. Together, the data collected in stages 1 and 2 (literature and interviews) allowed: 1) the systematization and proposition of a service taxonomy structure in three main categories (Infrastructure, Support and Education) that are subdivided into the subcategories 'Management and data curation' and 'Others'. These two subcategories have additional subdivisions that are replicated for each of the main categories. The services categorized as Support were the most frequent in the 167 documents (165 articles and two interviews) used to construct the taxonomy. Next are the services of the Educational category and, finally, the services of the Infrastructure category. The two most present categories - Support and Education - mainly comprise reference, instruction and guidance services, representing an extension of the traditional services already provided by libraries, now with a focus on research data; 2) the identification and compilation of the research data management support services offered; 3) the exposure of the challenges and difficulties faced by the interviewed institutions; and 4) the synthesis of the general actions for the viability and maintenance of services in the institutions. Step 2 (data collection through interviews) also allowed for a: 5) contextualization of the efforts for the development of RDMSSs in the interviewed institutions. We considered that the research met the proposed objectives of identifying services, actions, difficulties and challenges; contextualizing the offer of services and infrastructure in the interviewed universities; systematizing and proposing a taxonomy of services; synthesizing the difficulties and challenges; and synthesizing the actions to facilitate and maintain services in the institutions.

Keywords: Research data; Research data management; Data management support services; Libraries; Services.

LISTA DE FIGURAS

FIGURA 1 - ILUSTRAÇÃO ESQUEMÁTICA DOS DADOS DA CAUDA LONGA	37
FIGURA 2 - CENÁRIOS DE CICLO DE VIDA DOS DADOS	52
FIGURA 3 - DCC CURATION LIFECYCLE MODEL	55
FIGURA 4 - DDI COMBINED LIFE CYCLE MODEL	58
FIGURA 5- DATAONE DATA LIFE CYCLE	59
FIGURA 6 - USGS SCIENCE DATA LIFECYCLE MODEL	61
FIGURA 7 - RESEARCH DATA LIFECYCLE: UK DATA SERVICE	63
FIGURA 8 - LOCAIS IDEAIS DE SUBMISSÃO E CITAÇÃO COM O USO DOS PERIÓDICOS DE DADOS	72
FIGURA 9 - ÁREAS DE ENGAJAMENTO DAS BIBLIOTECAS NA GESTÃO DE DADOS DE PESQUISA	93
FIGURA 10 - PIRÂMIDE DA GESTÃO DE DADOS DE PESQUISA NAS BIBLIOTECAS	95
FIGURA 11 - ESTRUTURA DE GESTÃO DE DADOS CENTRADA NO USUÁRIO	99
FIGURAS 12 - COMPONENTES DO SERVIÇO DE SUPORTE À GESTÃO DE DADOS DE PESQUISA	113
FIGURA 13 – DESENHO DA PESQUISA	136
FIGURA 14 – CATEGORIAS E SUBCATEGORIAS	137
FIGURA 15 - FORMAÇÃO DOS NOMES DAS SUBCATEGORIAS	139
FIGURA 16 – INFRAESTRUTURA: CATEGORIAS E SUBCATEGORIAS DE SERVIÇOS	139
FIGURA 17 – SUPORTE: CATEGORIAS E SUBCATEGORIAS DE SERVIÇOS	140
FIGURA 18 – INFRAESTRUTURA: VISÃO GERAL DAS SUBCATEGORIAS	143
FIGURA 19 – SUPORTE: VISÃO GERAL DAS SUBCATEGORIAS	144
FIGURA 20 – EDUCACIONAL: VISÃO GERAL DAS SUBCATEGORIAS	145
FIGURA 21 - CATEGORIAS E SUBCATEGORIAS: NÚMERO DE ARTIGOS E REFERÊNCIAS AOS SERVIÇOS	147
FIGURA 22 – INFRAESTRUTURA: NÚMERO DE DOCUMENTOS E REFERÊNCIAS AOS SERVIÇOS	148
FIGURA 23 – INFRAESTRUTURA: SUBCATEGORIA GESTÃO E CURADORIA DE DADOS	149
FIGURA 24 – INFRAESTRUTURA: SUBCATEGORIA OUTROS	150
FIGURA 25 - SUPORTE – NÚMERO DE DOCUMENTOS E REFERÊNCIAS AOS SERVIÇOS	152
FIGURA 26 – SUPORTE: SUBCATEGORIA GESTÃO E CURADORIA DE DADOS	153
FIGURA 27 – SUPORTE: SUBCATEGORIA OUTROS	156
FIGURA 28 – EDUCACIONAL: NÚMERO DE DOCUMENTOS E REFERÊNCIAS AOS SERVIÇOS	158
FIGURA 29 – EDUCACIONAL: SUBCATEGORIA GESTÃO E CURADORIA DE DADOS	160
FIGURA 30 - EDUCACIONAL: SUBCATEGORIA OUTROS	162
FIGURA 31 - SUBCATEGORIA GESTÃO E CURADORIA DE DADOS: SERVIÇOS MAIS FREQUENTES	164
FIGURA 32 - SUBCATEGORIA OUTROS: SERVIÇOS MAIS FREQUENTES	165

LISTA DE QUADROS

QUADRO 1 - CATEGORIAS DE DADOS	25
QUADRO 2 – TIPOS DE DADOS	26
QUADRO 3 – DISTINÇÃO DOS DADOS	26
QUADRO 4 - TIPOLOGIA DOS DADOS	27
QUADRO 5 – ESCIENCE: VARIAÇÃO TERMINOLÓGICA ENTRE OS AUTORES	31
QUADRO 6 - CARACTERÍSTICAS DA <i>BIG SCIENCE</i> E DA <i>SMALL SCIENCE</i>	40
QUADRO 7 - ATIVIDADES TÍPICAS REALIZADAS NO CICLO DE VIDA DOS DADOS DE PESQUISA	53
QUADRO 8 - AÇÕES DO CICLO DE VIDA COMPLETO	56
QUADRO 9 - AÇÕES SEQUENCIAIS	56
QUADRO 10 - AÇÕES OCASIONAIS	57
QUADRO 11- BENEFÍCIOS DO COMPARTILHAMENTO DE DADOS	65
QUADRO 12 - CLASSIFICAÇÃO DE REPOSITÓRIOS	80
QUADRO 13 – ELEMENTOS DE UMA POLÍTICA PARA GESTÃO DE DADOS DE PESQUISA	85
QUADRO 14 - DENOMINAÇÃO DOS SERVIÇOS	109
QUADRO 15 - TIPOS DE SERVIÇOS OFERECIDOS OU PLANEJADOS PELAS BIBLIOTECAS	121
QUADRO 16 – INICIATIVAS DE GESTÃO DE DADOS NAS INSTITUIÇÕES DE ENSINO SUPERIOR BRASILEIRAS	131
QUADRO 17 – RESPOSTAS À CARTA-CONVITE E PEDIDO DE INFORMAÇÕES	132
QUADRO 18 – CATEGORIAS DE SERVIÇOS	137
QUADRO 19 – SUBCATEGORIA GESTÃO E CURADORIA DE DADOS	138
QUADRO 20 – SUBCATEGORIA OUTROS	138
QUADRO 21 – GESTÃO E CURADORIA DE DADOS: ASSUNTOS PRESENTES NAS SUBCATEGORIAS	141
QUADRO 22 – OUTROS: ASSUNTOS PRESENTES NAS SUBCATEGORIAS	142
QUADRO 23 – DESAFIOS E DIFICULDADES APONTADAS NA LITERATURA	181
QUADRO 24 - AÇÕES PARA VIABILIZAÇÃO E MANUTENÇÃO DE SSGDP NAS INSTITUIÇÕES	185

LISTA DE TABELAS

TABELA 1 – INFRAESTRUTURA: FREQUÊNCIA DE REFERÊNCIAS AOS SERVIÇOS	149
TABELA 2 – SUPORTE: FREQUÊNCIA DE REFERÊNCIAS AOS SERVIÇOS	152
TABELA 3 – SUPORTE: TOTAL DE REFERÊNCIAS NA CATEGORIA	154
TABELA 4 - EDUCACIONAL: FREQUÊNCIA DE REFERÊNCIAS AOS SERVIÇOS	158
TABELA 5 - SUBCATEGORIAS MAIS REFERENCIADAS	164

LISTA DE ABREVIATURAS E SIGLAS

ABNT	Associação Brasileira de Normas Técnicas
ACLS	<i>American Council of Learned Societies</i>
ALA	<i>American Library Association</i>
ARK	<i>Archival Resource Key</i>
ARL	<i>Association of Research Libraries</i>
ASIS&T	<i>American Society for Information Science & Technology</i> <i>Bases de Dados Referencial de Artigos de Periódicos em Ciência da Informação</i>
BRAPCI	<i>Collaborative Assessment of Research Data Infrastructures and Objectives</i>
CARDIO	
CEDAP	Centro de Documentação e Acervo Digital da Pesquisa
CERN	<i>European Organization for Nuclear Research</i>
CLOCKSS	<i>Controlled LOCKSS</i>
CNRI	<i>Corporation for National Research Initiatives</i>
CNSTAT	<i>Committee on National Statistics</i>
CODATA	<i>Committee on Data for Science and Technology</i>
CNPq	Conselho Nacional de Desenvolvimento Científico e Tecnológico
CPD	Centro de Processamento de Dados
CSU	<i>Colorado State University</i> <i>Consortium of Universities for the Advancement of Hydrologic Science</i>
CUAHSI	<i>Integrated Data Management System for Critical Zone Observatories</i>
CZOData	
DAF	<i>Data Asset Framework</i>
DataONE	<i>Data Observation Network for Earth</i>
DCC	<i>Digital Curation Center</i>
DDI	<i>Data Documentation Initiative</i>
DMWG	<i>Data Management Working Group</i>
DOI	<i>Digital Object Identifier</i>
E-IRG	<i>e-Infrastructure Reflection Group</i>
ESIP	<i>Earth Science Information Partners</i>
ESRC	<i>Economic and Social Research Council</i>
FAPEMIG	Fundação de Amparo à Pesquisa do Estado de Minas Gerais
FAPESP	Fundação de Amparo à Pesquisa do Estado de São Paulo
FOI	<i>Freedom of Information</i>
FURG	Universidade Federal do Rio Grande
GFBio	<i>German Federation for Biological Data</i>
IBICT	Instituto Brasileiro de Informação em Ciência e Tecnologia
IFLA	<i>International Federation of Library Associations</i>
KU	<i>Kansas University</i>
LISA	<i>Library and Information Science Abstracts</i>
LISTA	<i>Library, Information Science and Technology Abstracts</i>
LOCKSS	<i>Lots of Copies Keep Stuff Safe</i>
LSST	<i>Large Synoptic Survey Telescope</i>
NASEM	<i>National Academies of Sciences, Engineering, and Medicine</i>

NLM	<i>National Library of Medicine</i>
NNLM	<i>National Network of Libraries of Medicine</i>
NRC	<i>National Research Council</i>
NSB	<i>National Science Board</i>
NSF	<i>National Science Foundation</i>
OAI-PMH	<i>Open Archives Initiative - Protocol for Metadata Harvesting</i>
ORCID	<i>Open Researcher and Contributor ID</i>
OSU	<i>Oregon State University</i>
PROPESQ	Pró-Reitoria de Pesquisa da UFSC
PROPG	Pró-Reitoria de Pós-Graduação da UFSC
RCUK	<i>Research Councils UK</i>
RDM	<i>Research data management</i>
RNP	Rede Nacional de Ensino e Pesquisa
SEAD	<i>Sustainable Environmental Actionable Data</i> Superintendência de Governança Eletrônica e Tecnologia da
SeTIC	Informação e Comunicação
SIBI/USP	Sistema Integrado de Bibliotecas da Universidade de São Paulo
SINOVA	Secretaria de Inovação
SSGDP	Serviços de Suporte à Gestão de Dados de Pesquisa
TI	Tecnologia da informação
TIC	Tecnologia de informação e comunicação
UC	<i>University of California</i>
UFABC	Universidade Federal do ABC
UFMG	Universidade Federal de Minas Gerais
UFRGS	Universidade Federal do Rio Grande do Sul
UFSC	Universidade Federal de Santa Catarina
UFSCAR	Universidade Federal de São Carlos
UIUC	<i>University of Illinois Urbana-Champaign</i>
UNESP	Universidade Estadual Paulista
UNIFESP	Universidade Federal de São Paulo
URL	<i>Uniform Resource Locator</i>
USEPA	<i>United States Environmental Protection Agency</i>
USGS	<i>United States Geological Survey</i>
USP	<i>Universidade de São Paulo</i>
WWW	<i>World Wide Web</i>

SUMÁRIO

1.	INTRODUÇÃO	17
1.1.	<i>PROBLEMA DE PESQUISA</i>	20
2.2.	<i>OBJETIVOS</i>	21
2.	FUNDAMENTAÇÃO TEÓRICA	23
2.1.	<i>DADOS: UMA COMPREENSÃO NECESSÁRIA</i>	23
2.1.1.	Dados de pesquisa	24
2.2.	<i>O PARADIGMA DA eSCIENCE</i>	28
2.3.	<i>BIG SCIENCE E SMALL SCIENCE: CRIAÇÃO E ESTRUTURAÇÃO DOS DADOS</i>	33
2.3.1.	Dados e práticas de dados na cauda longa da ciência	38
2.4.	<i>GESTÃO DOS DADOS DE PESQUISA</i>	41
2.4.1.	Preservação e curadoria de dados de pesquisa	45
2.5.	<i>CICLO DE VIDA DOS DADOS</i>	48
2.5.1.	Modelos de Ciclo de vida dos dados	48
2.5.1.1.	DCC Curation Lifecycle Model	54
2.5.1.2.	DDI Combined Life Cycle Model	57
2.5.1.3.	DataONE Data Life Cycle	59
2.5.1.4.	USGS Science Data Lifecycle Model	60
2.5.1.5.	Research Data Lifecycle – UK Data Service	63
2.6.	<i>COMPARTILHAMENTO E REUSO DE DADOS</i>	64
2.7.	<i>PUBLICAÇÃO DE DADOS</i>	67
2.7.1.	Modelos de publicação de dados	70
2.8.	<i>CITAÇÃO DE DADOS</i>	73
2.9.	<i>INFRAESTRUTURAS, PADRÕES E POLÍTICAS PARA GESTÃO DE DADOS</i>	75
2.9.1.	Infraestrutura tecnológica	78
2.9.1.1.	Repositórios digitais	79
2.9.1.2.	Repositórios de dados de pesquisa	82
2.9.2.	Infraestrutura política	84
2.9.3.	Padrões de dados e metadados	86
2.10.	<i>AS BIBLIOTECAS E A GESTÃO DE DADOS DE PESQUISA</i>	89

2.10.1.	Biblioteconomia de dados	90
2.10.2.	Bibliotecas: ampliação de papéis	92
2.10.3.	Os bibliotecários e a gestão de dados de pesquisa: papéis e habilidades	97
2.11.	<i>PRÁTICAS E NECESSIDADES DOS PESQUISADORES</i>	101
2.12.	<i>SERVIÇOS DE SUPORTE À GESTÃO DE DADOS DE PESQUISA</i>	105
2.12.1.	Desenvolvimento de serviços: caminhos	110
2.12.2.	Elementos ou componentes dos SSGDPs	112
2.12.2.1.	Políticas e estratégias para gestão de dados de pesquisa	113
2.12.2.2.	Planos de negócio e sustentabilidade	114
2.12.2.3.	Orientação, treinamento e suporte	114
2.12.2.4.	Planejamento da gestão de dados	115
2.12.2.5.	Gestão de dados ativos	116
2.12.2.6.	Seleção e entrega de dados	116
2.12.2.7.	Repositório de dados	117
2.12.2.8.	Catálogo de dados	118
2.12.3.	Tipos e classificações de serviços	119
2.12.3.1.	Classificação dos serviços	121
2.12.4.	Desafios para o oferecimento de SSGDPs nas bibliotecas	123
3.	PROCEDIMENTOS METODOLÓGICOS	127
3.1.	<i>CARACTERIZAÇÃO DA PESQUISA</i>	127
3.2.	<i>ETAPAS DA PESQUISA</i>	128
3.2.1.	Primeira etapa: identificação de serviços e ações	128
3.2.2.	Segunda etapa: identificação e contextualização de serviços, ações e dificuldades	130
3.2.3.	Terceira etapa: análise dos dados	134
4.	APRESENTAÇÃO E DISCUSSÃO DOS RESULTADOS	137
4.1.	<i>TAXONOMIA DE SERVIÇOS: APRESENTAÇÃO E ESTRUTURA</i>	137
4.1.1.	Conteúdo das subcategorias	140
4.1.1.1.	Categoria Infraestrutura	143
4.1.1.2.	Categoria Suporte	143
4.1.1.3.	Categoria Educacional	144
4.2.	<i>TAXONOMIA DE SERVIÇOS: ANÁLISE DESCRITIVA E COMPILAÇÃO DE SSGDPs</i>	147
4.2.1.	Categoria Infraestrutura	147
4.2.2.	Categoria Suporte	151
4.2.3.	Categoria Educacional	157

4.2.4.	Subdivisões mais referenciadas	163
4.2.5.	Apontamentos gerais	166
4.3.	<i>CONTEXTUALIZAÇÃO NAS INSTITUIÇÕES ENTREVISTADAS</i>	167
4.3.1.	Gestão de dados de pesquisa na UFSC	168
4.3.2.	Gestão de dados de pesquisa na UFRGS	172
4.3.3.	Síntese e reflexões: entrevistas	177
4.4.	<i>DESAFIOS E DIFICULDADES</i>	181
4.5.	<i>AÇÕES PARA VIABILIZAÇÃO E MANUTENÇÃO</i>	184
5.	CONSIDERAÇÕES FINAIS	189
5.1.	<i>CONTRIBUIÇÕES DA PESQUISA</i>	195
5.2.	<i>ESTUDOS FUTUROS</i>	195
	REFERÊNCIAS	197
	APÊNDICE A – PADRÕES DE METADADOS	219
	APÊNDICE B – EXTENSÕES	221
	APÊNDICE C – LISTA DE ARTIGOS	223
	APÊNDICE D – ROTEIRO DE ENTREVISTA	230
	APÊNDICE E – CATEGORIA 1 - INFRAESTRUTURA: TRECHOS	231
	APÊNDICE F – CATEGORIA 2 - SUPORTE: TRECHOS	232
	APÊNDICE G – CATEGORIA 3 - EDUCACIONAL: TRECHOS	235
	APÊNDICE I – TAXONOMIA DE SERVIÇOS: NÚMERO DE REFERÊNCIAS	238
	APÊNDICE J – LISTA DE SERVIÇOS DA CATEGORIA 1 - INFRAESTRUTURA	240
	APÊNDICE K – LISTA DE SERVIÇOS DA CATEGORIA 2 - SUPORTE	242
	APÊNDICE L – LISTA DE SERVIÇOS DA CATEGORIA 3 - EDUCACIONAL	246

1. INTRODUÇÃO

A pesquisa científica tornou mais intensiva a geração e uso de dados (BORGMAN, 2012; SAYÃO; SALES, 2018), fenômeno observado a partir da *eScience*, nova forma de fazer pesquisa, apoiada numa infraestrutura altamente tecnológica e colaborativa. Apesar da eclosão na chamada *big science* – marcada por grandes projetos científicos em áreas como a Astronomia, Genômica e Física de Partículas –, a *eScience* engloba todas as áreas do conhecimento, abrangendo as milhares de pesquisas que produzem pequenos conjuntos de dados, porém heterogêneos, nas diversas disciplinas científicas da *small science*, também chamada de ‘cauda longa da ciência’¹ (*long tail of science*) (E-IRG, 2016; KRAFT, 2015). Na cauda longa, grande quantidade de pequenos projetos produzem pequenas quantidades de dados. Entretanto, se somadas, essas reduzidas quantidades de dados são maiores que aqueles gerados pela *big science* (SALES; SAYÃO, 2018).

Assim, pesquisas realizadas individualmente, pequenas equipes ou pequenos laboratórios, que formam a estrutura de pesquisa, principalmente, das universidades, também tornaram-se mais intensivos na geração e uso de dados (BORGMAN, 2012; SAYÃO; SALES, 2018).

Além disso, é observado reconhecimento do potencial informacional dos dados de pesquisa por parte de pesquisadores, instituições acadêmicas e agências de fomento. Os dados de pesquisa “[...] se devidamente tratados, preservados e gerenciados, podem constituir uma fonte inestimável de recursos informacionais para a pesquisa científica e para o ensino da ciência (SAYÃO; SALES, 2015, p. 1).

Dessa forma, a gestão de dados de pesquisa, que inclui todo o ciclo de vida dos dados desde o planejamento, coleta, análise, preservação e compartilhamento visando o reuso, tem ganhado destaque a partir das políticas mandatórias das agências financiadoras, sobretudo em âmbito internacional.

No Brasil, a Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) já “[...] reconhece a importância da gestão adequada dos dados de pesquisa como parte essencial das boas práticas de pesquisa” (FAPESP, 2019, *online*). Dessa forma, os dados resultantes de pesquisas financiadas pela FAPESP devem ser gerenciados e compartilhados. Os pesquisadores, ao submeterem projetos, devem obrigatoriamente incluir o Plano de Gestão de Dados (PGD): documento que descreve como os dados serão obtidos, gerenciados, compartilhados e preservados.

¹ A cauda longa da ciência é discutida na seção **2.3 Big science e small science: criação e estruturação dos dados**.

Em 2007 a *Association of Research Libraries* (ARL) já levantava questões importantes sobre os serviços e as infraestruturas nas bibliotecas para o apoio à pesquisa científica e gestão de dados, pois parte considerável do trabalho científico não está visível aos pares, nem à sociedade, sobretudo na cauda longa, segmento responsável pela maior parcela de dados produzidos pela ciência (SALES; SAYÃO, 2018).

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.

De outro lado, para as bibliotecas que até então estavam acostumadas a trabalhar com a pós-publicação, e que nesse momento, se veem obrigadas a oferecer serviços como apoio ao pesquisador, trabalhando com a pré-publicação desde o momento da criação dos dados, atuando, assim, na gestão de dados de pesquisa como um todo.

Dessa forma, as bibliotecas universitárias precisam assumir a responsabilidade de auxiliar na gestão e preservação de conjuntos de dados decorrentes de pesquisas realizadas dentro das universidades.

Algumas bibliotecas de instituições acadêmicas brasileiras já estão incluindo a gestão de dados de pesquisa como parte dos serviços oferecidos, como é o caso do Sistema Integrado de Bibliotecas da Universidade de São Paulo (SIBI/USP), onde os SSGDPs já estão estruturados e disponibilizados em seu *site*², e a Biblioteca Universitária da Universidade Federal de Santa Catarina, onde os serviços estão em fase de desenvolvimento. Durante participação no evento 'I Seminário de Suporte à Pesquisa e Gestão de Dados Científicos: panorama atual e desafios' realizado em setembro de 2017 pela Universidade Federal de Santa Catarina (UFSC), percebeu-se grande interesse na concepção e desenvolvimento desses serviços, principalmente nas bibliotecas acadêmicas do Estado de São Paulo, uma vez que já existe demanda a partir das exigências impostas aos pesquisadores pela FAPESP, agência financiadora daquele estado.

² <https://www.sibi.usp.br/apoio-pesquisador/dados-pesquisa>

Em Minas Gerais, pesquisa preliminar realizada no *site*³ da Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) não identificou qualquer menção acerca da gestão de dados. A FAPEMIG determina que, como condição básica para o recebimento de financiamento, os resultados das pesquisas financiadas devem, obrigatoriamente, “[...] ser publicados em periódicos indexados e especializados de circulação nacional ou internacional, quando de natureza científica e inovadora ou em mídias alternativas, quando de outra natureza” (FAPEMIG, 2019, *online*). No entanto, acredita-se que, a exemplo da FAPESP, as demais fundações estaduais de amparo à pesquisa devem caminhar para a exigência do plano de gestão de dados, como já vem acontecendo com as principais agências governamentais ou privadas de fomento à pesquisa da América do Norte, Europa e Austrália (FAPESP, 2017), impulsionado a gestão de dados de pesquisa gerados nas universidades de todo país.

A ideia de estudar essa temática surgiu a partir do trabalho desenvolvido pela autora como bibliotecária na Universidade Federal de Minas Gerais (UFMG) e pela percepção de um momento oportuno - grande valor atribuído aos dados de pesquisa e demandas das instituições de fomento. Isso já ocorre com frequência em nível internacional e está chegando ao Brasil neste momento.

Este trabalho trata da gestão de dados produzidos no âmbito científico, portanto, excluem-se os dados gerados por empresas, governo e sociedade.

Nesta tese, o conjunto de serviços visando apoio às atividades de gestão de dados de pesquisa é denominado Serviços de Suporte à Gestão de Dados de Pesquisa (SSGDPs). O estudo pretende propor uma taxonomia de serviços voltados à gestão de dados de pesquisa gerados nas universidades, bem como as ações necessárias para sua viabilização e manutenção. Além disso, o trabalho compila os diversos serviços oferecidos e identifica as principais dificuldades e desafios para o desenvolvimento e sustentabilidade dos SSGDPs nas instituições.

A tese está organizada da seguinte forma: o primeiro capítulo apresenta a contextualização da pesquisa, problemas e objetivos. O capítulo 2 apresenta a fundamentação teórica da pesquisa: os dados de pesquisa, o paradigma da *eScience*, *big science* e *small science*, gestão de dados de pesquisa, ciclo de vida dos dados, compartilhamento e reuso de dados científicos, publicação e citação de dados, as infraestruturas, padrões e políticas, as bibliotecas e a gestão de dados de pesquisa, incluindo os desafios enfrentados para a entrega de serviços de apoio à gestão de dados pelas bibliotecas, as práticas e

³ <http://www.fapemig.br/>

necessidades dos pesquisadores e os Serviços de Suporte à Gestão de Dados de Pesquisa (SSGDPs). O capítulo 3 apresenta os procedimentos metodológicos da pesquisa. O capítulo 4 apresenta e discute os resultados e, finalmente, o capítulo 5 apresenta as considerações finais deste trabalho.

1.1. PROBLEMA DE PESQUISA

Os pesquisadores realizam pesquisas científicas e produzem grande quantidade de dados em todas as áreas do conhecimento, englobando tanto o segmento da *big science* quanto da cauda longa da ciência. Esses pesquisadores precisam gerenciar e disponibilizar seus dados de pesquisa com vistas à reutilização, necessidade surgida, principalmente, a partir das políticas mandatórias das agências financiadoras.

Para que os dados sejam disponibilizados para uso futuro e, efetivamente, reutilizados, a instituição precisa fornecer, aos pesquisadores, infraestruturas como repositórios, políticas e padrões que norteiam a gestão de dados. Além disso, precisam oferecer assistência aos cientistas em diversas questões relacionadas à gestão de dados de pesquisa como: plano de gestão de dados; inclusão de metadados; uso de repositórios; armazenamento; preservação, dentre outros.

Dessa maneira, o suporte à gestão dos dados gerados dentro das universidades, produzidos pelos grandes projetos científicos ou pelas investigações realizadas na esfera da cauda longa da ciência, torna-se desafio para as bibliotecas. Essas deverão assumir papéis diferentes daqueles tradicionais, tornando-se parte do ciclo da pesquisa, papel até então negligenciado por essas instituições, criando serviços que atendam às mudanças ocorridas no cenário da comunicação científica, como a valorização dos dados de pesquisa, que passaram a figurar como produtos de ‘primeira classe’ da mesma forma que os artigos científicos publicados em periódicos. Os seus serviços deverão focar ativamente o apoio ao pesquisador para a realização da gestão adequada de dados em todo o ciclo de vida da pesquisa, desde o planejamento até seu reuso.

Às bibliotecas cabem agora, dentre outros, serviços como: ajuda/consultoria na elaboração do plano de gestão de dados; definição de padrões de metadados; identificação de repositórios confiáveis; suporte de referência para encontrar e citar dados; e realização de treinamentos voltados para a gestão de dados oriundos de pesquisas realizadas dentro da universidade.

Assim, identifica-se como problema de pesquisa os desafios da gestão dos dados a serem enfrentados pelas bibliotecas universitárias, para que possam oferecer suporte aos pesquisadores através de seus serviços.

A partir do problema de pesquisa pergunta-se: quais serviços as bibliotecas poderão oferecer e como poderão oferecer? Quais as infraestruturas necessárias, as ações que deverão ser executadas e as dificuldades e desafios enfrentados para o oferecimento dos Serviços de Suporte a Gestão de Dados de Pesquisa nas instituições?

Parte-se do pressuposto que as práticas, tipos de dados gerados/coletados e as necessidades dos pesquisadores; opiniões e expectativas dos gestores institucionais; liderança da biblioteca; e habilidades dos profissionais envolvidos influenciam nos tipos de serviços que podem ser oferecidos pelas bibliotecas, conforme explicitado na literatura científica da área.

2.2. OBJETIVOS

Objetivo geral

O objetivo deste estudo é propor uma taxonomia de serviços voltados à gestão de dados de pesquisa gerados nas universidades, bem como sintetizar dificuldades, desafios e ações necessárias para a viabilização e manutenção desses serviços nas instituições.

Objetivos específicos

1. Identificar, por meio de revisão de literatura:
 - a) os serviços voltados à gestão de dados de pesquisa promovidos por bibliotecas acadêmicas em âmbito mundial;
 - b) as ações necessárias para viabilização e manutenção desses serviços;
 - c) as dificuldades e desafios em relação à implantação dos serviços de suporte à gestão de dados nas instituições.
2. Identificar e contextualizar, por meio da visão dos atores institucionais:
 - a) os serviços voltados à gestão de dados de pesquisa oferecidos pelas instituições;
 - b) as ações necessárias para viabilização e manutenção desses serviços;
 - c) as dificuldades e desafios em relação à implantação dos serviços de suporte à gestão de dados de pesquisa nas instituições.
3. Sistematizar e propor uma taxonomia de serviços de suporte à gestão de dados a partir da revisão de literatura e visão dos atores institucionais.

4. Sintetizar as dificuldades e desafios em relação à implantação dos serviços de suporte à gestão de dados de pesquisa nas instituições, conforme visão da literatura e atores institucionais.
5. Sintetizar, por meio da revisão de literatura e visão dos atores institucionais, as ações gerais para a implantação e manutenção dos serviços de suporte à gestão de dados de pesquisa nas instituições.

2. FUNDAMENTAÇÃO TEÓRICA

Este capítulo apresenta a fundamentação teórica da tese, fornecendo contexto para o desenvolvimento e empenho das bibliotecas com a gestão de dados de pesquisa. Assim, são abordados, dentre outros assuntos, os dados, a gestão de dados de pesquisa, compartilhamento e reuso de dados, publicação e citação, as bibliotecas e a gestão de dados, as infraestruturas, padrões e políticas e os SSGDPs.

2.1. DADOS: UMA COMPREENSÃO NECESSÁRIA

A explosão informacional ocorrida pós II Guerra Mundial provocou grande interesse pelas atividades de ciência e tecnologia e, conseqüentemente, crescimento exponencial do conhecimento registrado. Vivemos, atualmente, a segunda grande transformação, onde se produz enorme quantidade de dados nas atividades de pesquisa científica (CORRÊA, 2017).

Os dados podem ser considerados a força vital da ciência e tecnologia, desempenhando papel importante na criação, teste, revisão de hipóteses, teorias e modelos (FREDERICK, 2016). Os dados ainda podem ser vistos como insumos para a criação da informação e do conhecimento. Corrêa (2017) complementa que os dados são o combustível para o avanço científico.

Os dados tem se feito presentes nas discussões e literatura científica:

Todo mundo está falando sobre dados. Cientistas experimentais vivem e respiram dados. Os teóricos são desafiados por dados. Os financiadores estão se perguntando como tornar mais acessíveis os dados produzidos com seu suporte sem esticar seus orçamentos. As comunidades de pesquisa estão buscando novos depósitos de dados e padrões para apoiá-los. E os editores científicos estão se perguntando como hospedar dados e fornecer controle de qualidade (WELCOME..., 2014, p. 534, tradução nossa).

Os dados, contudo, podem assumir diversas formas, sendo seu conceito de difícil definição (BORGMAN, 2012). O *National Research Council* (NRC) (1999, p. 15, tradução nossa), braço operacional da *The National Academies of Sciences, Engineering, and Medicine* (NASEM) dos Estados Unidos define dados como “[...] fatos, números, letras e símbolos que descrevem um objeto, ideia, condição, situação ou outros fatores”. Assim, o dado é a menor unidade de informação existente.

Para a *National Science Foundation* (NSF) (2007, p. 22, tradução nossa) a definição de dados é mais ampla, englobando todos e quaisquer dados gerados “[...] a partir de observações, experimentos, simulações, modelos e montagens de ordem superior [...]”, além de toda “[...] documentação associada necessária para descrever e interpretar os dados”.

No relatório publicado pelo conselho americano *National Science Board* (NSB) (NSB, 2005, p. 13, tradução nossa) o termo ‘dados’ refere-se a qualquer informação armazenada em formato digital “[...] incluindo textos, números, imagens, vídeo ou filmes, áudio, *software*, algoritmos, equações, animações, modelos, simulações *etc.*” gerados de várias formas como observação, computação ou experimento.

Dessa maneira, na maioria das vezes, os dados são pensados como “[...] uma coleção de dígitos, cadeias de texto ou códigos alfanuméricos que podem não parecer ter algum significado inerente” (FREDERICK, 2016, p. 11, tradução nossa), representando algo que ocorreu ou está ocorrendo no mundo real.

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. Auckland (2012) faz distinção entre dados e informação. Dados referem-se a itens discretos como imagens e números, sem significado intrínseco, ao passo que informação são dados interpretados e analisados.

2.1.1. Dados de pesquisa

Especificamente falando sobre os dados gerados pelas atividades de pesquisa, os chamados ‘dados de pesquisa’, Corrêa (2017) afirma que esses compreendem uma área muito ampla, pois estão presentes em todas as disciplinas. Buscar uma definição é tarefa desafiadora, pois tal definição pode variar em relação à abordagem dos diferentes atores (pesquisadores, instituições, financiadores) e em relação aos diversos contextos nacionais no qual a pesquisa está inserida.

Contudo, de forma geral, Corrêa (2017, p. 434, tradução nossa) define dados de pesquisa como “[...] informação registrada ou produzida mediante qualquer forma ou meio durante o transcurso de uma pesquisa” podendo ser do tipo numéricos, descritivos ou visual, em formato papel (notas de pesquisa em cadernos, fotografias *etc.*) ou digital. Os dados de pesquisa podem incluir os protocolos, códigos numéricos, gráficos e tabelas necessárias para coletar e organizar dados, tanto em trabalho de campo quanto laboratório. Dessa maneira, os dados de pesquisa incluem, também, os resumos gerados durante o transcurso de realização de uma pesquisa.

Corrêa (2017, p. 434, tradução nossa) complementa com outra definição: “os dados de pesquisa são todas as evidências que um pesquisador necessita para validar suas conclusões após uma pesquisa”, sendo resultantes da investigação acadêmica em todas as áreas do conhecimento (KIM, 2013).

Por serem produzidos ou recolhidos de diferentes formas, de acordo com Willis, Greenberg e White (2012, p. 1506, tradução nossa) os dados de pesquisa científica, ou dados científicos “[...] variam muito, assim como o seu uso pretendido e potencial ao longo do tempo”. Devido a essa variedade e com o objetivo de “[...] ser útil para entender as semelhanças e diferenças, bem como o uso pretendido e potencial dos dados ao longo do tempo”, eles podem ser classificados ou agrupados de várias maneiras, seja pela disciplina que cria e planeja usar os dados ou pelo método de coleta de dados.

Diversas são as classificações e tipologias de dados. Salieta-se que todas as distinções implicam escolhas que devem ser feitas para arquivamento e preservação (NSB, 2005). O relatório do *Committee on Data for Science and Technology* (CODATA) (1974) desenvolveu uma classificação para os dados de pesquisa no âmbito da ciência e tecnologia. A classificação consiste em 15 tipos de dados divididos em seis categorias, conforme Quadro 1 a seguir:

QUADRO 1 - Categorias de dados

Categorias de dados	Tipos de dados	Descrição
Dados independentes e dependentes do tempo	Dados independentes do tempo	Dados que podem ser medidos repetidamente
	Dados dependentes do tempo	Dados que podem ser medidos apenas uma vez
Dados de localização independentes e dependentes	Localização independente	Dados independentes da localização dos objetos medidos
	Dependente da localização	Dados dependentes da localização de objetos medidos
Dados primários, derivados e teóricos	Dados observacionais ou experimentais primários	Dados obtidos por experiência ou observação
	Dados derivados	Dados derivados pela combinação de vários dados primários com a ajuda de um modelo teórico
	Dados teóricos	Dados obtidos por cálculos teóricos
Dados determináveis e estocásticos	Dados determináveis	Dados sobre uma quantidade que pode assumir um valor definido sob uma determinada condição
	Dados estocásticos	Dados sobre uma quantidade que leva valores flutuantes de uma amostra ou medida para outro
Dados quantitativos e qualitativos	Dados quantitativos	Medidas de quantidades científicas em unidades bem definidas
	Dados semiquantitativos	Medidas de quantidades científicas usando escalas arbitrárias
	Dados qualitativos	Qualquer declaração científica definitiva sobre objetos científicos
Dados como valores numéricos, gráficos ou modelos	Dados como valores numéricos	Dados apresentados como valores numéricos isolados
	Dados como modelos ou gráficos	Dados apresentados em forma gráfica ou como modelos
	Dados simbólicos	Dados apresentados usando símbolos arbitrários (não-numéricos)

Fonte: Elaborado pela autora com base em CODATA (1974) e Willis, Greenberg e White (2012).

O *National Research Council* (1999) divide os dados em: brutos ou não tratados, processados e verificados (Quadro 2).

QUADRO 2 – Tipos de dados

Tipos	Descrição
Brutos	Dados obtidos a partir de observações originais, como os dados coletados por satélite e transmitidos à Terra, ou resultados experimentais iniciais, como dados de testes laboratoriais.
Processados	Dados que após a coleta passaram por processamento ou refinamento de forma a torná-los mais utilizáveis, ordenados ou simplificados, aumentando sua inteligibilidade.
Verificados	São os dados cuja qualidade e precisão foram asseguradas. Nos dados experimentais a verificação significa que os dados mostraram ser reproduzíveis em um teste ou experiência que repete o original. Em dados observacionais a verificação significa que os dados foram comparados com outros dados cuja qualidade é conhecida ou que o instrumento com o qual foram obtidos foi devidamente calibrado e testado.

Fonte: Elaborado pela autora com base em NRC (1999).

Para o *National Science Board* (2005), os dados podem ser distinguidos por sua natureza diversificada, sua origem e pelo nível de processamento a que foram submetidos (Quadro 3).

QUADRO 3 – Distinção dos dados

Distinção	Descrição
Natureza	Diversificada - números, imagens, fluxos de vídeo ou áudio, informações de versão de <i>software</i> e <i>hardware</i> , algoritmos, equações, animações ou modelos/simulações.
Origem (reprodutibilidade)	Observacional - dados obtidos, por exemplo, a partir de observações diretas da temperatura do oceano em uma data específica; ou as atitudes dos eleitores antes de uma eleição; ou fotografias de um evento astronômico, um registro histórico que não pode ser lembrado.
	Computacional - dados resultantes de um modelo ou simulação computacional.
	Experimental - dados como as medições de padrões de expressão gênica; taxas de reação química; ou desempenho do motor.
Derivação (nível de processamento)	Por processamento ou pelas atividades de curadoria. Quando a partir da análise e refinamento dos dados brutos são geradas diferentes versões.

Fonte: Elaborado pela autora com base em NSB (2005).

Para Corrêa (2017) a primeira distinção básica entre os tipos de dados é que eles podem ser classificados em quantitativos e qualitativos, contudo essa classificação não é tão evidente, já que pesquisadores de todas as áreas utilizam os dois tipos. Segundo Corrêa (2017), a *National Science Foundation* (NSF) classifica os dados segundo os critérios apresentados no quadro a seguir (Quadro 4):

QUADRO 4 - Tipologia dos dados

Crítérios	Tipos	Descrição
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 preservadas a documentação de <i>hardware</i> e <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 <i>etc.</i> ; 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.
	Semi-estruturados	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 produtos nas mesmas condições.
Formato	Arquivos digitais	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 costumes de cada disciplina; grande quantidade de formatos disponíveis; opção mais segura para garantir o acesso; e o uso de formatos padrões.

Fonte: Elaborado pela autora com base em Corrêa (2017).

Cada uma dessas diferenças implica decisões políticas importantes em relação à preservação dos dados. Por exemplo,

Os dados observacionais não podem ser recolhidos e são arquivados indefinidamente. Os dados que são o resultado de modelos de computador ou simulações podem ser reproduzidos se forem fornecidas informações adequadas sobre *hardware*, *software* e insumos do computador. Geralmente, os dados experimentais podem ser reproduzidos, embora existam casos em que as condições ou variáveis experimentais são desconhecidas. (WILLIS; GREENBERG; WHITE, 2012, p. 1506, tradução nossa).

A criação de dados nas áreas do conhecimento acontece de formas bem variadas (BORGMAN, 2012). Cada disciplina ou subdisciplina tem seu próprio conjunto de características de dados. O tamanho e o tipo do projeto determinarão a heterogeneidade e o nível de complexidade desses dados (FRIEDLANDER; ADLER, 2006). A compreensão do que significam os dados (e suas classificações) vai depender do contexto em que os dados foram criados. Podem advir de um tipo específico de pesquisa, disciplinas e de necessidades diferentes dos pesquisadores (SWANSON; RINEHART, 2016).

O uso de novos métodos e instrumentação faz com que as pesquisas científicas gerem enorme quantidade de dados (BORGMAN, 2012). Para Henty (2008) a expansão da capacidade das tecnologias de informação e comunicação (TICs) fez com que pesquisadores inventassem novas formas de fazer trabalho aproveitando os benefícios dessas tecnologias para apoio aos cálculos, análises, modelagem, simulação, colaboração e publicação, fazendo com que, em seus processos de pesquisa, os cientistas coletem e criem dados massivamente. Essa nova forma de fazer investigação é denominada *eScience*, uma ciência orientada por dados. Assim, a próxima seção apresenta o paradigma da *eScience*.

2.2. O PARADIGMA DA eSCIENCE

A *internet* mudou a forma como os cientistas trabalham (THORPE, 2009), principalmente, como organizam e trocam informações, transformando a vida dos pesquisadores (SEMELER; PINTO; VIANNA, 2017). Outra mudança importante foi na quantidade de dados que os cientistas acumulam e precisam analisar, além do modo de armazenar e recuperá-los (THORPE, 2009).

Desde o início da década de 1990, a *web* tem sido ferramenta indispensável no mundo científico para a troca de informação. Cerca de uma década após as grandes mudanças promovidas pelo surgimento da WWW⁴, os cientistas viram-se obrigados a desenvolver

⁴ *World Wide Web*

capacidades de colaboração que vão além da *web*, impulsionados pelo dilúvio de dados – *petabytes*⁵ de dados - gerados a partir de experimentos de alto rendimento, simulações de supercomputadores e demais pesquisas científicas (HEY; TREFETHEN, 2005). Além de acessar informações de diferentes *sites*, os cientistas “[...] querem usar recursos de computação remota, integrar, federar⁶ e analisar informações de muitos recursos de dados distribuídos e díspares, e acessar e controlar equipamentos experimentais remotos (HEY; HEY, 2006, p. 516-517, tradução nossa), sendo a capacidade de acessar, mover, manipular e extrair dados requisito central das novas aplicações de ciência colaborativa, criando um ambiente propício ao desenvolvimento da *eScience*.

Para Sayão e Sales (2017a), o surgimento dessa nova forma de fazer ciência, chamada *eScience*

[...] desloca a pesquisa científica contemporânea na direção de uma ciência orientada por dados, em que o maior problema não é a escassez, mas sim o excesso de dados e a capacidade de interpretar seus padrões ocultos na forma de conhecimento e novas descobertas (SAYÃO; SALES, 2017a, p. 29).

Dessa forma, os procedimentos e práticas tradicionais dos pesquisadores passaram e ainda passam por mudanças radicais. Essas alterações na forma de condução da ciência são descritas por uma série de termos como ciberciência (*cyberscience*), ciberinfraestrutura (*cyberinfrastructure*) e *e-Science* (JANKOWSKI, 2007).

O primeiro deles, a ciberciência, seria, de acordo com Nentwich (2003, p. 22, tradução nossa), “[...] todas as atividades de pesquisa acadêmica e científica no espaço virtual gerado pelos computadores em rede e pelas tecnologias avançadas de informação e comunicação [...]”. O autor complementa: a ciberciência “[...] é o que os pesquisadores fazem no ciberespaço [...]”, um espaço virtual criado pelas redes eletrônicas, sendo mais que apenas as formas eletrônicas de fazer ciência. Ciberciência não é o estudo do ciberespaço, mas ciência ou pesquisa realizada no ciberespaço ou nas condições do ciberespaço, ao contrário da ciência tradicional, que não usa computadores em redes.

Similar ao termo infraestrutura, usado desde 1920 para designar as estradas, redes elétricas, pontes e demais obras públicas necessárias ao funcionamento de uma economia industrial, a ciberinfraestrutura, definida em 2003 pela *National Science Foundation* (NSF, 2003) seria uma “[...] infraestrutura baseada em computação distribuída e tecnologia da

⁵ *Petabyte* (PB) - 1 PB = 1.000.000.000.000.000 Bytes

⁶ De Sistemas de Informação Federados – integração de dados provenientes de sistemas heterogêneos e independentes, porém cooperativos. Esses sistemas possibilitam uma busca federada, realizada em vários recursos simultaneamente (ALMEIDA; CENDÓN, 2019).

informação e comunicação” necessária para uma economia do conhecimento, envolvendo elementos de *hardware*, *software*, dados e pessoal (NSF, 2003, p. 5, tradução nossa).

Já o termo *eScience* foi cunhado em 1999 por John Taylor, então diretor-geral dos *Research Councils* (RCUK) do Reino Unido (UK). Ele reconheceu a importância da era da *eScience* e garantiu que o Reino Unido continuasse a desempenhar papel de liderança na transformação científica. Os departamentos de computação britânicos foram os pioneiros no que agora chamamos de *eScience*. Outras iniciativas semelhantes surgiram na mesma época em outros países, como os projetos de ciberinfraestrutura nos Estados Unidos (THORPE, 2009).

O termo *eScience*, de acordo com Brandt (2007, p. 365, tradução nossa) também é sinônimo de *e-Research*: “[...] uma rede mais ampla que abrange todos os domínios de pesquisa, desafiados pela exploração de novas formas de fazer pesquisas em ambientes computacionalmente aprimorados e em rede”.

Apesar de Lewis (2010) considerar o termo *e-Research* mais inclusivo para as disciplinas ditas não científicas que estão, cada vez mais, usando as técnicas e ferramentas da *eScience*, neste trabalho, optamos por utilizar o termo *eScience*. De forma não exaustiva, no quadro a seguir (Quadro 5) elencamos alguns termos usados.

QUADRO 5 – eScience: variação terminológica entre os autores

Termo	Autor (ano)
<i>Cyberinfrastructure</i>	(NSF, 2007; GOLD, 2007a, b; HEY; TREFETHEN, 2005; JANKOWSKI, 2007; NRC, 2011; NSF, 2003)
<i>Cyberscience</i>	(JANKOWSKI, 2007; NENTWICH, 2003)
<i>Data-driven science</i>	(CESAR JUNIOR, 2011)
<i>Data-intensive computing</i>	(CESAR JUNIOR, 2011)
<i>Data-Intensive Research</i>	(TENOPIR <i>et al.</i> , 2015; TENOPIR; BIRCH; ALLARD, 2012)
<i>Data-rich, data-driven science</i>	(GOLD, 2007a)
<i>Digitally enhanced science</i>	(FOX; HENDLER, 2011; OSSWALD, 2008)
<i>e-infrastructure</i>	(HEY; TREFETHEN, 2005)
<i>Enhanced science</i>	(JANKOWSKI, 2007)
<i>eResearch</i>	(HALL; DE ROURE; SHADBOLT, 2009; HENTY, 2008; LUCE, 2008)
<i>e-Research</i>	(BRANDT, 2007; HENTY, 2008; PATERSON <i>et al.</i> , 2007; STEWARD; LEWIS, 2010)
<i>eScience</i>	(CESAR JUNIOR, 2011; HAYES, 2014; HEY; TANSLEY; TOLLE, 2011; THORPE, 2009; WRIGHT <i>et al.</i> , 2007)
<i>e-Science</i>	(ARL, 2007; CASACUBERTA; VALLVERDÚ, 2014; DAUREEN, 2008; GOLD, 2007a; HENTY, 2008; HEY; TREFETHEN, 2003, 2002, 2005; HINE, 2006; JANKOWSKI, 2007; OSSWALD, 2008; SEMELER; PINTO; VIANNA, 2017; SZIGETI; WHEELER, 2011; VERETENNIKOVA <i>et al.</i> , 2015; WATSON; TREFETHEN; VANDER MEER, 2010)
<i>Fourth paradigm</i>	(HEY; TANSLEY; TOLLE, 2011; TENOPIR; BIRCH; ALLARD, 2012)
<i>From data to knowledge</i>	(CESAR JUNIOR, 2011)
<i>Networked, data-driven science</i>	(HEY; HEY, 2006)
<i>Research 2.0</i>	(KOLTAY, 2016a, 2017; KOLTAY; ŠPIRANEC; KARVALICS, 2015)
<i>Science data</i>	(POOLE, 2015)

Fonte: Elaborado pela autora.

Notamos que vários termos são usados para se referir ao fenômeno da *eScience*, sendo possível perceber grande variação terminológica entre os autores.

A *eScience*, contudo, não é uma nova disciplina científica (HEY; TREFETHEN, 2005) (HEY; HEY, 2006, p. 517, tradução nossa), mas “[...] uma forma abreviada para o conjunto de ferramentas e tecnologias necessárias para apoiar a ciência colaborativa e em rede”(HEY; HEY, 2006, p. 517, tradução nossa).

Para Semeler, Pinto e Vianna (2017) a *eScience* é vista como fenômeno da informação contemporânea, ou, como novo fenômeno informativo que usa novas técnicas para organizar e trocar informações científicas por meio do uso de métodos de investigação baseados em infraestrutura de *hardware* e *software*.

Dessa maneira, Sayão e Sales (2017a) afirmam que a *eScience* se desenvolve no ambiente da ciberinfraestrutura, altamente tecnológico, onde a geração de conhecimento é realizada cada vez mais por equipes colaborativas, multidisciplinares e distribuídas em escala mundial.

Na *eScience*, o uso de novas tecnologias de informação e colaboração científica, como recursos para modelagem e análise de dados, cadernos de laboratório eletrônicos e/ou digitalizados, conjuntos de dados brutos e ajustados, versões de projetos preliminares (manuscritos, pré-impressões e/ou publicações eletrônicas) e manipulação da própria informação em formato digital, permite que cientistas conduzam as pesquisas de forma mais eficiente e colaborativa, alterando a forma como as pesquisas são realizadas na ciência tradicional (SEMELER; PINTO; VIANNA, 2017).

Semeler, Pinto e Vianna (2017) apontam que as metodologias de investigação mudaram a partir das novas infraestruturas tecnológicas. As metodologias da ciência tradicional como observação/exploração empírica, teorização e os últimos estudos sobre simulação são ampliados quando mediados por essas infraestruturas tecnológicas contemporâneas, proporcionando, diferentes metodologias e técnicas para coleta de dados e gerando nova dimensão para os cientistas e nova forma de fazer ciência.

Os conceitos de *eScience* emergiram primeiramente nas ciências – medicina, tecnologia e engenharia. Pesquisadores destas áreas já caminhavam para projetos mais interdisciplinares e colaborativos devido ao alto custo de alguns equipamentos e a enorme quantidade de dados gerados – na casa dos *petabytes* – que ainda hoje, só podem ser analisados por supercomputadores. O uso de instrumentos dispendiosos por cientistas é facilitado pelo estabelecimento de uma ciberinfraestrutura de pesquisa (DAUREEN, 2008).

Apesar do surgimento, em primeiro momento, nas ciências físicas e naturais, a *American Council of Learned Societies (ACLS) Commission on Cyberinfrastructure for the Humanities*

and Social Sciences – afirma que pesquisadores das áreas de humanidades e ciências sociais também abraçaram a ideia da ciberinfraestrutura (DAUREEN, 2008).

Já a força tarefa conjunta da ARL complementa que a

eScience inclui todas as ciências naturais e físicas, disciplinas aplicadas e tecnológicas relacionadas, bem como biomedicina e ciências sociais que compartilham abordagens de pesquisa com as ciências naturais. (ARL, 2007, p. 4, tradução nossa).

Merz (2006, p. 100, tradução nossa), no entanto, alerta que a ideia de que a *eScience* é empreendimento coerente, que abrange igualmente ampla gama de campos científicos, é construção retórica e que “[...] não reflete adequadamente a natureza complexa e multifacetada da prática científica e cultural em uma era de informatização generalizada”. A autora defende a existência de diferentes culturas epistêmicas que incorporam, configuram e evoluem com as tecnologias de informação e comunicação (TICs) de formas diferenciadas, configurando a chamada desunião da *eScience* que pressupõe desenvolvimento desigual de diferentes campos científicos no que diz respeito à sua adoção e uso de infraestruturas digitais.

Merz (2006) afirma que as áreas do conhecimento se relacionam com a *eScience* de formas diferentes. Por consequência, a criação de dados nas áreas do conhecimento também acontece de formas bem variadas (BORGMAN, 2012). Nas áreas da *big science*, como Astronomia e Física de Partículas, a *eScience* é mais nitidamente notada. Contudo, a *eScience* também está presente nas áreas da *small science*, *little science* ou ainda *long tail science*, em pesquisas realizadas individualmente ou em pequenos laboratórios, a partir de novos métodos e instrumentação, que coletam e observam grandes volumes de dados (BORGMAN, 2012).

A discussão sobre as características e formas de criação e estruturação dos dados nas áreas da *big science* e *small science* serão tratadas a seguir.

2.3. BIG SCIENCE E SMALL SCIENCE: CRIAÇÃO E ESTRUTURAÇÃO DOS DADOS

Os termos *little science* e *big science* já são utilizados desde a década de 1950 para caracterizar e contrastar os estilos da organização do trabalho científico (BORGMAN *et al.*, 2015). Pouco depois, em 1963, Derek J. de Solla Price publicou o livro intitulado ‘*Little Science, Big Science*’ discutindo a transformação da estrutura da ciência, numa transição gradual da *little science* para a *big science* (PRICE, 1986). Borgman, Wallis e Enyedy (2007) acreditam que esses campos estão em transição à medida que a ciência se torna mais instrumentalizada.

Diferentes termos são utilizados para designar *big science* e *small science*. A *big science* (BICARREGUI *et al.*, 2013; BORGMAN, 2010, 2012, 2015; BORGMAN; WALLIS; ENYEDY, 2007; CARLSON, 2006; DAVENHALL, 2011; FERGUSON *et al.*, 2014; HEIDORN, 2008; MACMILLAN, 2014; PRICE, 1986; SAYÃO; SALES, 2013) é também conhecida como *modern science* (BORGMAN, 2015; PRICE, 1986) ou ainda *large science* (HEIDORN, 2008).

A *small science* (BORGMAN; WALLIS; ENYEDY, 2007; CARLSON, 2006; CRAGIN *et al.*, 2010; FERGUSON *et al.*, 2014; HEIDORN, 2008; LYON, 2012; MACMILLAN, 2014; ONSRUD; CAMPBELL, 2007) é também chamada de *little science* (BORGMAN *et al.*, 2015; BORGMAN; WALLIS; ENYEDY, 2007; PRICE, 1986), *small group and single investigator science* (ONSRUD; CAMPBELL, 2007) ou *long tail science* (HEIDORN, 2008; PALMER *et al.*, 2007). Esse último, traduzido por cauda longa da ciência (SAYÃO; SALES, 2018).

Borgman *et al.* (2015) argumentam que essa divisão da ciência continua a ser um paradigma influente dentro de várias disciplinas. No entanto, essas distinções entre *small* e *big science* são também problemáticas, pois tentam reduzir toda a complexidade da ciência em uma simples dicotomia. Esses autores preferem trabalhar como a escala em suas variadas dimensões (tamanho da equipe, financiamento, prazo, volume, tipo e forma dos dados gerados) atua na prática de dados em vez de trabalhar apenas com a distinção entre *small science* e *big science*.

Segundo Borgman (2015), Price⁷ distinguiu as ciências não pelo tamanho dos projetos, mas pela maturidade científica das áreas. A *big science* “[...] é caracterizada por esforços internacionais, colaborativos e por colégios invisíveis de pesquisadores que se conhecem e que trocam informações de forma formal e informal” (BORGMAN, 2015, p. 5, tradução nossa).

A *big science* é marcada pelo uso de grandes dispositivos e instrumentos de pesquisa; alto custo; longa duração de programas e projetos; grandes equipes; distribuição em escala planetária e com uma grande *expertise* no domínio na qual estão inseridos (SAYÃO; SALES, 2018).

A *small science* “[...] é conduzida em menor escala, com comunidades menores, menos acordo em questões e métodos de pesquisa e menos infraestrutura” sendo “tipificada por métodos heterogêneos, dados heterogêneos e por controle e análise locais” (BORGMAN, 2015, p. 5-6, tradução nossa).

⁷ Price, Derek John de Solla. *Little Science, Big Science*. New York: Columbia University Press, 1963.

Para Sayão e Sales (2018), as atividades da *big science* têm correlação com o fenômeno do *big data* – o *big data* científico. Os termos *big data* e *little data* são análogos à *big science* e *little science* (BORGMAN, 2015) e às suas variações terminológicas.

Os *big data* “[...] são caracterizados como grandes em volume, produzidos de forma contínua e de natureza variada, embora muitas vezes sejam subproduto de sistemas, ao invés de serem projetados para investigar fenômenos ou processos específicos (KITCHIN; LAURIAULT, 2015, p. 463, tradução nossa).

De acordo com Lynch (2008, p. 28, tradução nossa), os dados científicos podem ser considerados *big data* de várias formas. Primeiramente, pela grandeza e “[...] pela maneira como eles desafiarão o estado da arte em computação, rede e armazenamento de dados” principalmente originados a partir de projetos de pesquisa nacionais e internacionais como os projetos na área de física de partículas do *European Organization for Nuclear Research* (CERN) ou do *Large Synoptic Survey Telescope* (LSST). Os dados podem ser considerados *big data* por terem “[...] uma significância duradoura - um resultado de ensaio clínico ou a observação de um evento único” e pelos “[...] desafios descritivos que podem exigir contexto, como a configuração experimental”. Dessa forma, esses dados podem ser facilmente compartilhados, replicados e recombinaados, apresentando oportunidades de reutilização, permitindo acelerar as pesquisas em andamento e aproveitar os já realizados investimentos em pesquisa.

Os “*small data*” são caracterizados por seu volume geralmente limitado, coleção não contínua, variedade estreita e geralmente são gerados para responder a perguntas específicas” (KITCHIN; LAURIAULT, 2015, p. 463, tradução nossa). São produzidos ou coletados por indivíduos ou pequenos grupos e direcionados a abordar questões científicas específicas e muitas vezes com cobertura geográfica ou temporal limitada. Geralmente são insubstituíveis, caros de reproduzir, raramente reutilizados (ELAG *et al.*, 2017).

A distinção entre *big data* e *small data* é recente. Antes de 2008 todos os dados eram considerados pequenos independentemente do seu volume, devido ao “[...] custo, recursos e dificuldades de gerar, processar, analisar e armazenar dados [...]” (KITCHIN; LAURIAULT, 2015, p. 464, tradução nossa).

Essa distinção é problemática devido às diversas maneiras pelas quais algo pode ser grande. Os *big data* não têm tamanho absoluto, mas sua grandeza é questão de escala. “Os dados são grandes ou pequenos em termos do que pode ser feito com eles, quais os *insights* que eles podem revelar e a escala de análise necessária em relação ao fenômeno do interesse - seja o comportamento de compra do consumidor ou a descoberta da droga”

(BORGMAN, 2015, p. 6, tradução nossa). Contudo, a definição dos 3 V's⁸ continua a ser útil: volume, variedade, velocidade ou combinação destes. Dessa forma, “um aumento substancial em qualquer uma dessas dimensões dos dados pode levar a mudanças na escala de pesquisas [...]” (BORGMAN, 2015, p. 6, tradução nossa). A autora ainda complementa que a ubiquidade dos dados também contribui para a sua grandeza.

O rápido crescimento e impacto dos *big data* gerou alguns questionamentos: se esses poderiam levar à queda do *small data* ou se os estudos com base em *small data* poderiam ser diminuídos. Entretanto, os autores concluem que os *small data* continuarão a ser componente vital no cenário de pesquisa (KITCHIN; LAURIAULT, 2015). Os autores apresentam três argumentos que apoiam a ideia de que, apesar do rápido crescimento dos *big data*, bem como de suas análises, os pequenos estudos que geram *small data* continuarão a existir:

- a) o histórico dos *small data* em responder perguntas específicas;
- b) os *small data* serão cada vez mais agrupados, vinculados e escalados por meio de infraestruturas, dotados de padrões, formatos, metadados e documentação, aumentando seu valor de combinação e compartilhamento;
- c) os *small data* serão cada vez mais expostos às novas epistemologias da ciência dos dados e à incorporação em outras aplicações para o qual eles nunca foram planejados.

Apesar de os *small data* serem vitais no cenário de pesquisa (KITCHIN; LAURIAULT, 2015), muita atenção tem sido dada aos extensivos conjuntos de dados produzidos pelos grandes projetos científicos - *big data* - , em detrimento das milhares de pesquisas que produzem pequenos conjuntos de dados, porém heterogêneos, nas diversas disciplinas científicas da *small science* – as disciplinas da chamada cauda longa da ciência (*long tail of science*). (E-IRG, 2016; KRAFT, 2015).

A metáfora da cauda longa (*long tail*) é usada para caracterizar a disponibilidade e o uso de dados (BORGMAN, 2015) ou ainda, analisar a distribuição da geração de dados (SAYÃO; SALES, 2018) pela *big science* e *small science*.

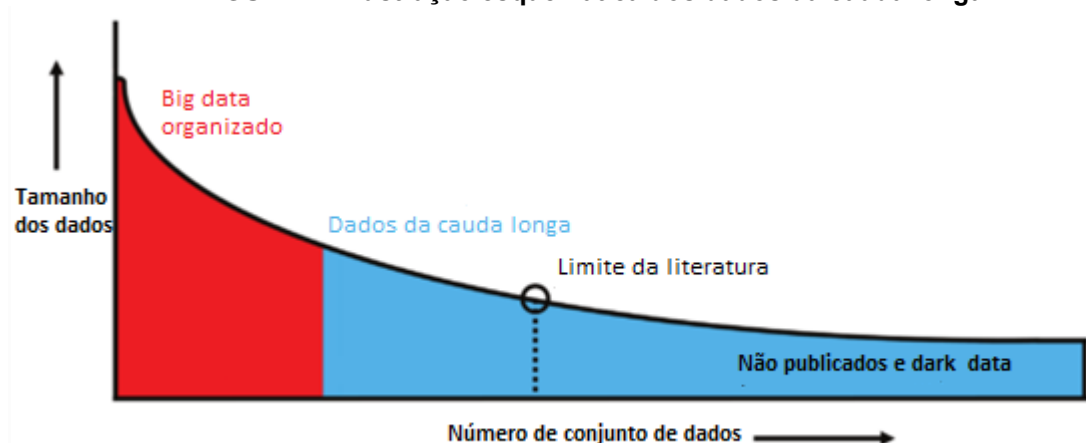
A metáfora é baseada no conceito estatístico da ‘cauda longa’ (*long tail*) (SAYÃO; SALES, 2018). O termo foi cunhado em 2004 por Chris Anderson em um artigo da Wired Magazine (BORGMAN, 2015), sendo muito utilizado nas áreas de *marketing*.

⁸ A definição de *big data* pode ser articulada em três dimensões ou três V's (volume, variedade e velocidade) sendo consideradas como características e desafios ao lidar com grandes dados (MANYIKA *et al.*, 2011, p. 1, tradução nossa).

No modelo de Anderson, cerca de 15% da distribuição está na cabeça da curva; os 85% restantes dos casos são distribuídos ao longo da cauda. Quando aplicados a pesquisas acadêmicas, um pequeno número de equipes de pesquisa trabalha com volumes de dados muito grandes, algumas equipes trabalham com dados muito pequenos e a maioria está em algum lugar intermediário. Na extrema direita da curva, um grande número de estudiosos conduz suas pesquisas com quantidades mínimas de dados (BORGMAN, 2015, p. 8, tradução nossa).

Dessa forma, uma pequena parte – 15% das pesquisas - produz uma grande quantidade de dados (*big data*), enquanto a maioria – 85% das pesquisas - produz poucos dados. Esses 85% das pesquisas estão distribuídas nas disciplinas da chamada cauda longa da ciência (*long tail of science*).

FIGURA 1 - Ilustração esquemática dos dados da cauda longa



Fonte: (FERGUSON *et al.*, 2014, p. 1443, tradução nossa).

A FIG. 1 mostra que os *big data* gerados pelos “[...] grandes esforços científicos bem organizados, com dados homogêneos e bem organizados, representam apenas pequena proporção dos dados coletados pelos cientistas” ao passo que “uma proporção muito grande de dados científicos cai no prolongamento da distribuição, com inúmeros pequenos esforços de pesquisa independentes que produzem grande variedade de conjuntos de dados de pesquisa especializada” (FERGUSON *et al.*, 2014, p. 1443).

Os autores apontam que a extrema direita da distribuição inclui os dados não publicados (*dark data*) que, apesar da riqueza, geralmente são inacessíveis. Embora possam ser resultados positivos ou negativos e existir tanto na *big science* quanto na *small science*, grande parte dos *dark data* reside na cauda longa da ciência, em inúmeros projetos e laboratórios. Assim, esses conjuntos de dados invisíveis que residem nos cadernos e discos rígidos dos pesquisadores (O’SHEA, 2008) são dificilmente encontrados por seus usuários potenciais, pouco reutilizados ou preservados (HEIDORN, 2008). Contudo, os *dark data* podem ser muito valiosos, constituindo-se desafio tornar esses dados úteis para a sociedade (HEIDORN, 2008).

A ciência da cauda longa não é sinônima de ciência pequena ou de questões científicas pequenas ou irrelevantes (HEIDORN, 2008). Apesar de o mundo ser cada vez mais dominado pelo *big data* (BERMAN, 2013), a grande quantidade de múltiplos projetos da cauda longa pode contribuir para grandes realizações, sendo terreno fértil para novas ideias, uma vez que é provável que a inovação na ciência venha da cauda longa (HEIDORN, 2008).

2.3.1. Dados e práticas de dados na cauda longa da ciência

Os dados, bem como as práticas de dados, diferenciam-se de acordo com a escala de pesquisa, em termos de instalações, requisitos de tecnologia, tamanho das equipes, tamanho e volume e diversidade de dados, financiamentos e outros fatores (BORGMAN *et al.*, 2015).

À medida que os projetos aumentam em escala, as condições em que os dados são coletados, armazenados, gerenciados, curados⁹ e tornados acessíveis tendem a tornar-se mais padronizados e rotineiros. Projetos menores geralmente são mais adaptáveis às condições locais (BORGMAN *et al.*, 2015, p. 210, tradução nossa).

Na cauda longa, o uso de padrões e melhores práticas para gestão de dados são extremamente variáveis (KRAFT, 2015). As coleções são altamente heterogêneas e tendem a ficar isoladas em escritórios e laboratórios de cientistas (PALMER *et al.*, 2007). Borgman *et al.* (2016) estudaram as características e as necessidades de gestão de dados dos laboratórios de pequeno e médio porte e descobriram que as equipes de pesquisa lutam com o volume de dados e com a falta de conhecimentos humanos e infraestruturas de conhecimento suficientes para gerenciar seus dados.

As práticas de dados nos laboratórios de pequeno e médio porte são diversas, mas muitos requisitos são comuns como: reunir, descrever, limpar, analisar dados e publicar resultados. Todos esses laboratórios apresentam uma variedade de problemas de gestão de dados que inclui falta de recursos, *expertise*, metadados e padronização (BORGMAN *et al.*, 2016).

Todos os pequenos e médios laboratórios estudados foram restringidos pela falta de conhecimentos técnicos em suas equipes, suporte técnico de suas instituições e disponibilidade geral de recursos humanos. Os seus dados foram espalhados por vários computadores, plataformas e ferramentas adquiridas durante longos períodos de tempo. Muitas vezes, as ferramentas e plataformas eram incompatíveis, tanto em seus laboratórios quanto entre seus laboratórios e seus colaboradores. Da mesma forma, seus dados e suas práticas foram altamente diversas devido à ampla gama de *expertise* disciplinar e ampla gama de questões de pesquisa em estudo (BORGMAN *et al.*, 2016, p. 141, tradução nossa).

Os laboratórios da ciência de cauda longa obtêm financiamento de várias fontes. Cada subvenção pode “[...] ter duração de três anos ou menos e apoiar um ou dois investigadores,

⁹ A discussão sobre a preservação e curadoria de dados é apresentada na seção 2.4.1.

além de ser conduzida por um pequeno grupo de estudantes e de bolsistas pós-doutorado. Esses laboratórios podem funcionar de forma independente ou fazer parte de colaborações locais ou distribuídas maiores” (BORGMAN *et al.*, 2016, p. 129, tradução nossa).

A coleta de dados e as ferramentas são muitas vezes adaptadas às condições científicas. Apesar de serem eficazes para a tarefa em questão, essas práticas muitas vezes resultam em dados difíceis de serem disseminados ou compartilhados (BORGMAN *et al.*, 2016).

Os laboratórios de pequeno e médio porte enfrentam frequentemente o problema do ‘atrito científico’ - “[...] dificuldades encontradas quando duas disciplinas científicas trabalhando em problemas relacionados tentam interoperar” (EDWARDS *et al.*, 2011, p. 669, tradução nossa) -, impedindo o fluxo de informações dentro e entre equipes de pesquisa. Isso dificulta o gerenciamento, a descrição e o armazenamento de dados. Esses atritos incluem a falta de padronização, metadados incompletos ou conflitantes, reivindicações de propriedade intelectual, práticas de compartilhamento de dados e restrições de recursos humanos (BORGMAN *et al.*, 2016).

Em pesquisas realizadas por pesquisadores individuais ou por pequenos laboratórios, os próprios pesquisadores são responsáveis pelo gerenciamento dos dados e, geralmente, usam formatos de dados locais (BORGMAN *et al.*, 2016). Muitas vezes, as pequenas equipes confiam em terceiros, por exemplo, estudantes e bolsistas, para gerenciar seus dados (BORGMAN *et al.*, 2016; BORGMAN; WALLIS; ENYEDY, 2007). Dessa forma, esses dados podem, às vezes, ser negligenciados e perdidos (BORGMAN *et al.*, 2016).

O quadro 6, a seguir, sintetiza as características da *big science* e *small science* em termos de dados e práticas.

QUADRO 6 - Características da *big science* e da *small science*

Características	<i>Big Science (Cabeça)</i>	<i>Small Science (cauda)</i>
Projetos	Projetos de grande escala, rotineiros e de longa duração	Projetos de escala menor, ocasionais e de pequena duração
	Instrumentação ampla	Instrumentação reduzida
	Alto custo	Baixo custo
	Grandes equipes (número de colaboradores) internacionais	Indivíduos ou pequenas equipes
	Pesquisa distribuída	Pesquisa local
	Ciclos de financiamento maiores (10 anos)	Ciclos de financiamento curtos (3 anos ou menos), menores do que o período em que os dados permanecem cientificamente úteis
Coleta e/ou produção de dados	Padrões e procedimentos comuns (métodos, ferramentas e técnicas)	Diversidade de padrões e procedimentos (métodos, ferramentas e técnicas)
	Rapidez na criação (em tempo real ou próximo a isso)	Criação mais lenta
	Mecanizado	Gerados à mão
Dados	Homogêneos (tipo e formato)	Heterogêneos (tipo e formato)
	Altamente estruturados	Não estruturados
	Grandes quantidades de dados (volume)	Pequeno ou limitado (volume)
	Abrange normalmente uma população inteira	Abrange amostras de uma população
	Forte relacionalidade (união a diferentes conjuntos de dados)	Fraca relacionalidade
	Alta flexibilidade de extensão (adicionar novos campos com facilidade) e escalabilidade (expansão de tamanho)	Baixa flexibilidade e escalabilidade
Gestão e curadoria	Gestão e curadoria central (bibliotecas digitais compartilhadas)	Gestão e curadoria local - o próprio pesquisador
	Mais recursos dedicados à gestão de dados	Poucos recursos para adquirir, curar, compartilhar e manter o acesso aos dados
	Repositórios disciplinares e de referência	Repositório institucional, geral ou ausência de repositório
	Preservado	Não preservado
Compartilhamento e reuso	Acesso aberto	Acesso obscurecido ou protegido
	Redes formais	Redes informais
	Imediatamente reusado	Raramente reusado

Fonte: Elaborado pela autora com base em Borgman (2012, 2015); Borgman *et al.* (2015, 2016); Borgman, Wallis e Enyedy (2007); Cragin *et al.* (2010); E-IRG (2016); Ferguson *et al.* (2014); Heidorn (2008); Kitchin e Lauriault (2015); Onsrud e Campbell (2007); Palmer *et al.* (2007); Wallis, Rolando e Borgman (2013).

Assim como na *big science*, gestão e curadoria de dados também são necessárias para agregar valor às coleções de dados produzidas pela *long tail science* e para facilitar sua integração com outras coleções de dados (PALMER *et al.*, 2007). Os autores acreditam que é necessária uma combinação de soluções de gestão de dados, desde as coleções grandes

e disciplinares com dados relativamente homogêneos até as coleções heterogêneas interinstitucionais, centralizadas ou distribuídas produzidas pela ciência da cauda longa. A gestão de dados de pesquisa é apresentada na próxima seção.

2.4. GESTÃO DOS DADOS DE PESQUISA

Atualmente, o uso generalizado de tecnologia computacional em todas as disciplinas fez com que pesquisadores gerem e usem grande quantidade de dados que são intrínsecos ao processo de pesquisa. Assim, “[...] armazenar estes dados de forma fácil de acessar, processar e analisar é um desafio” (COX; PINFIELD, 2014, p. 299, tradução nossa).

As pesquisas geram dados que precisam ser organizados e entendidos (CORRÊA, 2017). Contudo, lidar com esses dados é um dos grandes desafios do século XXI. Se os dados de pesquisa estiverem perdidos no computador, eles podem se tornar irrelevantes ou uma série de dígitos indecifráveis (CARLSON, 2006).

Nos últimos 10 anos surgiu uma agenda internacional para a gestão de dados de pesquisa (COX *et al.*, 2017; LATHAM, 2017) que compreende “[...] a organização dos dados, desde a sua entrada no ciclo de pesquisa até a divulgação e arquivamento de resultados valiosos” (WHYTE; JONATHAN, 2011, n.p., tradução nossa).

Vários conceitos e/ou atividades estão relacionadas à gestão de dados de pesquisa. Koltay (2016a) aponta que a gestão

significa cuidar dos dados de pesquisa, facilitar o acesso a eles, preservar e agregar valor a esses dados ao longo de seu ciclo de vida, que é um dos principais determinantes da publicação de dados abertos e da curadoria de dados, trazendo os benefícios de permitir encontrar e compreender dados, evitando duplicações desnecessárias e validando resultados (KOLTAY, 2016a, p. 97, tradução nossa).

Para Flores *et al.* (2015), a gestão de dados de pesquisa inclui a documentação, a curadoria e a preservação dos dados de pesquisa, assegurando em longo prazo o valor e a utilidade dos dados de pesquisa, possibilitando novas análises e replicação dos achados do estudo original.

Possivelmente, as definições de Flores *et al.* (2015) e Koltay (2016a) aproximam-se muito do conceito de curadoria, que será apresentado na seção seguinte. Entretanto, esses autores referem-se às atividades como gestão de dados de pesquisa.

Cox e Pinfield (2014) ampliam o conceito e argumentam que a gestão de dados de pesquisa

consiste em várias atividades e processos associados ao ciclo de vida dos dados, envolvendo *design* e criação de dados, armazenamento, segurança, preservação, recuperação, compartilhamento e reutilização, considerando

todas as capacidades técnicas, considerações éticas, questões legais e estruturas de governança (COX; PINFIELD, 2014, p. 300, tradução nossa).

Príncipe (2013) menciona outras atividades envolvidas na gestão de dados de pesquisa: planejamento e gestão de dados, criação, documentação, acesso e utilização, armazenamento e *backup*, compartilhamento e preservação de dados. Já Corrêa (2017) afirma que a gestão de dados de pesquisa inclui processos de preservação, uso e reuso de dados. Finalizando, Sayão e Sales (2015) afirmam que

a gestão de dados cobre todos os aspectos relativos à manipulação, organização, documentação e agregação de valor, e tem um papel crucial como facilitador nos processos de compartilhamento dos dados, na garantia da sustentabilidade e acessibilidade dos dados em longo prazo (SAYÃO; SALES, 2015, p. 1).

A gestão de dados de pesquisa foi impulsionada por uma série de fatores como a *eScience*, a demanda dos financiadores de pesquisa e editores e pelo reconhecimento da importância da gestão de dados de pesquisa para a melhoria do acesso aos resultados das investigações realizadas pelas instituições (COX *et al.*, 2017).

Além desses, Lyon (2012) aponta outros fatores que estimularam a gestão de dados de pesquisa como uma agenda de comunicação acadêmica cada vez mais aberta; a adoção de ferramentas e aplicações da *web* que aceleram o processo de publicação; e pressões no sentido de maior responsabilização e transparência para mostrar o impacto dos investimentos públicos na ciência.

Latham (2017), entretanto, afirma que o interesse pela gestão de dados de pesquisa foi impulsionado por questões puramente financeiras. Os patrocinadores de pesquisa começaram a exigir que os dados de pesquisa fossem gerenciados com o intuito de facilitar o acesso e o compartilhamento de dados, a validação e a reprodutibilidade da pesquisa.

Os fatores elencados anteriormente contribuíram para o crescente foco na gestão, uso, compartilhamento e reuso de dados científicos, tornando-se um desafio para as instituições (HENTY, 2008). Para que sejam reutilizados, os dados devem estar bem preservados. As agências financiadoras de pesquisa começaram a exigir de pesquisadores e instituições a documentação e implementação de planos de gestão e compartilhamento de dados, incluindo todo o ciclo de vida – do início até após a conclusão da pesquisa – uma atividade desafiadora que inclui questões de responsabilidade, financiamento, recursos e a decisão de quais dados devem ser mantidos e por quanto tempo (LYNCH, 2008).

Além da comunidade de pesquisa e das agências financiadoras, os governos também estão interessados na gestão de dados científicos:

[...] os governos estão interessados em apoiar a pesquisa e, ao mesmo tempo, desejosos de assegurar que os produtos de pesquisa, ou seja, os dados sejam bem gerenciados, facilmente acessíveis e, de preferência, disponíveis em acesso aberto (HENTY, 2008, p. 1, tradução nossa).

Discutindo sobre o porquê da gestão de dados, Lewis (2010) afirma que a resposta é banal: porque o volume de dados gerados pelas pesquisas é enorme e, ainda pelas exigências das agências de financiamento. Além disso, a gestão de dados representa potencial benefício para a pesquisa acadêmica em si, pois incentiva o compartilhamento de dados de pesquisa.

Corroborando Lewis (2010), Tenopir *et al.* (2017) afirmam que uma boa gestão de dados é especialmente importante visando à garantia da transparência da pesquisa científica, a preservação dos dados, à possibilidade de reuso e reanálise e ao avanço do conhecimento, exercendo papel importante para a instituição no fornecimento de infraestrutura de suporte à *eScience* (HENTY, 2008). Corrêa (2017) considera que a gestão adequada dos dados é primordial, pois pode permitir aos cientistas novas formas de comparação e descobertas e a geração de novos campos de investigação.

Apesar de primordial, diferentemente do desenvolvimento que a pesquisa conseguiu alcançar, as responsabilidades pela gestão de dados e o suporte a todas as questões envolvidas ainda não alcançaram o mesmo patamar. Henty (2008) argumenta que

o ciclo das comunicações acadêmicas não incorporou até agora a importância da gestão de dados como parte desse ciclo. Isso está mudando à medida que os dados são vistos como resultado de pesquisa de igual importância para as publicações (HENTY, 2008, p. 2, tradução nossa).

Além disso, as instituições precisam decidir sobre quem recai a responsabilidade sobre os dados (HENTY, 2008).

Para Lynch (2008), os desafios da gestão de dados somente serão resolvidos com a colaboração entre instituições, financiadores e cientistas. Todos os atores envolvidos na infraestrutura de pesquisa e comunicação científica estão em busca de seu papel e responsabilidades junto à gestão de dados (GOLD, 2007a).

Em nível institucional, Márdero Arellano (2017) destaca que há uma lacuna na preparação das instituições para lidar com grandes volumes de dados. A criação de políticas que envolva a gestão dos dados de pesquisa é importante em todo o seu ciclo, fomentando a conscientização sobre essas questões.

No campo das políticas, o primeiro passo poderia ser atrelar a gestão e o compartilhamento de dados como obrigação para as pesquisas com financiamento governamental. Outra necessidade incluiria a criação de padrões e práticas para citação de dados. O terceiro grande passo seria o desenvolvimento de práticas e mecanismos de controle de acesso

adequado aos dados, pois em casos onde exista limitação de acesso ou restrições quanto à propriedade intelectual. É importante salientar que iniciativas dessa natureza já são encontradas em âmbito mundial (GOLD, 2007a).

Assim como as instituições, as agências de financiamento devem elaborar uma política de dados que estipule as obrigações que o pesquisador deve cumprir como condição do financiamento (MÁRDERO ARELLANO, 2017).

Para Márdero Arellano (2017), os pesquisadores são os principais responsáveis pelos dados científicos, uma vez que eles têm o contato mais direto com a criação ou coleta de dados. Dessa forma, são eles que entendem como usar os dados. Aos pesquisadores cabe o planejamento da coleta, curadoria, descrição e disseminação dos dados no início da pesquisa. Tudo isso deve ser informado no plano de gestão de dados (PGD). Já para Corrêa (2017) os pesquisadores podem apoiar esse processo de duas formas:

- a) organizando os dados de maneira adequada em relação ao seu ciclo de vida - planejamento desde a coleta de dados até o armazenamento seguro garantindo a continuidade de novas pesquisas;
- b) depositando os dados em repositórios e revistas com estrutura adequada para indexação e recuperação.

Dentre os diversos atores envolvidos ou interessados na gestão de dados de pesquisa estão as bibliotecas, que sempre ofereceram serviços voltados para as pesquisas. Elas se tornaram importantes protagonistas, à medida que enormes quantidades de dados digitais se tornaram preponderantes, surgindo necessidade de gerenciá-los de forma mais urgente. As bibliotecas começaram a incorporar a gestão de dados nos seus serviços (FLORES *et al.*, 2015) – chamados neste trabalho de Serviços de Suporte à Gestão de Dados de Pesquisa (SSGDPs) e que serão discutidos em seções posteriores.

Percebemos que a gestão de dados pode ter significados diferentes e nem sempre pode ser diferenciada da curadoria de dados. Koltay (2016b) confirma essa percepção ao apontar que os conceitos de gestão e curadoria de dados não parecem estar claramente separados.

Apesar de os conceitos de gestão de dados de pesquisa e curadoria estarem bastante imbricados, acreditamos que a curadoria e a preservação de dados são componentes da gestão de dados de pesquisa, conforme apresentado por Márdero Arellano (2017), Sayão (2017) e Príncipe (2013), e que engloba outros componentes como infraestruturas, políticas, questões legais, éticas e propriedade intelectual, seleção e coleção, descrição, organização, acesso, compartilhamento, reuso, segurança, custos, capacitação, dentre outros.

A seção seguinte aborda os aspectos da preservação e curadoria digital dos dados de pesquisa, consideradas neste trabalho como uma das atividades ou ações da gestão de dados científicos.

2.4.1. Preservação e curadoria de dados de pesquisa

A gestão de recursos digitais envolve a preservação e curadoria. Friedlander e Adler (2006), citando Charles Humphrey, ex-coordenador do *Research Data Services* da *University of Alberta* no Canadá, afirma que os conceitos de preservação – origens nos arquivos - e curadoria – origem nos museus - foram emprestados de outras áreas e aplicados ao domínio dos dados de pesquisa.

A preservação é “[...] uma atividade de arquivamento na qual itens específicos de dados são mantidos ao longo do tempo para que possam ser acessados e compreendidos através de sucessivas mudanças e obsolescência de tecnologias”. Já o arquivamento é uma das atividades da curadoria “[...] que garante que os dados sejam devidamente selecionados, armazenados, possam ser acessados e que sua integridade lógica e física seja mantida ao longo do tempo, incluindo segurança e autenticidade” (LORD; MACDONALD, 2004, p. 3, tradução nossa).

A preservação abarca o uso de práticas de gestão baseadas em padrões de dados e metadados ao longo de todo ciclo de vida da pesquisa e que orienta os cuidados de longo prazo a serem tomados com os dados digitais, em relação ao acesso, migração *etc.* (FRIEDLANDER; ADLER, 2006).

Para o *Data Thesaurus*, disponibilizado pelo site NNLM RD3: *Resources for Data-Driven Discovery*, a preservação de dados

[...] consiste em uma série de atividades [...] necessárias para garantir o acesso contínuo aos dados pelo tempo necessário. A preservação de dados requer gerenciamento ativo contínuo de dados, o mais cedo possível, no ciclo de vida. À medida que os volumes, a heterogeneidade e a complexidade dos dados aumentam, esse requisito para o gerenciamento ativo se torna componente muito importante (NNLM, 2018, *online*, tradução nossa).

No gerenciamento ativo e preservação, deve-se garantir que os dados da pesquisa sejam “[...] organizados e descritos com metadados apropriados, armazenados em local onde possam ser monitorados e disponibilizados para uso futuro e preparados para migração ou transformação, caso o formato de dados esteja danificado obsoleto” (NNLM, 2018, *online*, tradução nossa).

A preservação é um dos principais desafios da curadoria (JOINT, 2007), pois, “[...] se os dados precisarem ser armazenados, eles deverão ser movidos de plataforma para plataforma, pois as plataformas decairão e serão substituídas” (2007, p. 453, tradução nossa).

A expressão curadoria de dados (*data curation*) foi cunhada por Diane Zorich, em 1995, a partir de experiências com objetos do museu e informação de espécimes. A autora afirma que “o objeto do museu e a informação do espécime são dinâmicos e cumulativos, e grande parte dessa informação resulta dos processos que esses materiais passam no museu” (ZORICH, 1995, p. 430, tradução nossa). Dessa forma, por analogia, “[...] o termo foi estendido também aos objetos digitais, transferindo essas abordagens para as coleções digitais” (SALES, 2014, p. 54).

Tradicionalmente, a curadoria não se preocupa apenas com o cuidado em longo prazo de livros, pinturas ou outros artefatos. Também procura manter sua integridade, permitir e promover sua disponibilidade para o público apropriado. O mesmo cenário é verdadeiro para os dados (LORD; MACDONALD, 2004, p. 3, tradução nossa).

A curadoria também cuida da manutenção e aprimoramento dos dados, para que estes permaneçam sempre úteis (LORD; MACDONALD, 2004). Assim, de maneira similar à curadoria exercida pelos museus, os dados também podem ser curados (ZORICH, 1995). A curadoria é vista como

atividade de gerenciar e promover o uso de dados a partir de seu ponto de criação, para garantir que [os dados] estejam aptos para uso contemporâneo e disponível para descoberta e reutilização. Para conjuntos de dados dinâmicos, isso pode significar enriquecimento ou atualização contínuos para mantê-los aptos para o propósito (LORD; MACDONALD, 2004, p. 3, tradução nossa).

E envolve

[...] maneiras de organizar, exibir e redirecionar coleções de dados preservadas. [...] As funções de curadoria agregam valor a uma coleção de dados preservados, organizando e exibindo os dados por meio de análises dos metadados da coleção ou da criação de novos dados da coleção preservada (FRIEDLANDER; ADLER, 2006, p. 18, tradução nossa).

Para o *Digital Curation Center* (DCC) do Reino Unido, a curadoria digital inclui manter, preservar e agregar valor aos dados de pesquisa ao longo de seu ciclo de vida (DCC, 2017).

O *Data Thesaurus* (NNLM, 2018), assim como Abbott (2010), entende a curadoria digital como atividades envolvidas na gestão de dados, desde o planejamento da sua criação até a disponibilidade e adequação para descoberta e reuso no futuro. A curadoria digital é

[.] a gestão e preservação de dados digitais a longo prazo. Todas as atividades envolvidas na gestão de dados desde o planejamento de sua

criação, implementação de melhores práticas em digitalização e documentação, e garantia de sua segurança, disponibilidade e adequação para descoberta e reutilização no futuro, fazem parte da curadoria digital. A curadoria digital pode incluir o gerenciamento de conjuntos de dados para uso diário, por exemplo, garantindo que eles possam ser pesquisados e continuar legíveis com o tempo (NNLM, 2018, *online*, tradução nossa).

Dessa maneira, a curadoria digital abrange uma gestão ativa sobre os dados desde sua criação, pensando também no futuro: na sua preservação para uso e reuso:

[...] a curadoria digital envolve a gestão atuante e a preservação de recursos digitais durante todo o ciclo de vida de interesse do mundo acadêmico e científico, tendo como perspectiva o desafio temporal de atender a gerações atuais e futuras de usuários (SAYÃO; SALES, 2012, p. 184).

O principal desafio da curadoria digital de dados de pesquisa “[...] recai na necessidade de se preservar não somente o conjunto de dados, mas de preservar, sobretudo, a capacidade que ele possui de transmitir conhecimento para uso futuro das comunidades interessadas” (SAYÃO; SALES, 2012, p. 180).

Assim, entende-se que a curadoria digital é conceito abrangente (POOLE, 2015) ou ainda paradigma (YAKEL, 2007) que envolve a preservação digital (YAKEL, 2007) e a curadoria de dados (POOLE, 2015; YAKEL, 2007), que se refere à forma como os dados são cuidados (RODRIGUES *et al.*, 2010) e pressupõe um conjunto de técnicas ou ações que envolvem a gestão e a preservação, de forma a manter os dados e outros materiais digitais durante o seu ciclo de vida, conferindo valor imediato e garantindo sustentabilidade para todas as gerações de usuários (BEAGRIE, 2006; SALES, 2014).

No âmbito da curadoria de dados de pesquisa, observamos que as expressões ‘curadoria digital’, ‘curadoria de dados’, ‘curadoria de dados de pesquisa’, ‘curadoria digital de dados de pesquisa’ ou simplesmente ‘curadoria’ são indistintamente usadas, porém com os mesmos propósitos. Consideramos mais adequada a expressão ‘curadoria digital de dados de pesquisa’ por abranger maior grau de especificidade para designar as ações de curadoria voltadas especificamente aos dados de pesquisa em formato digital.

As ações de curadoria não terminam com o fim dos projetos, pois os dados evoluem. É necessária uma gestão contínua (SAYÃO; SALES, 2017b). A gestão - incluindo a curadoria de dados de pesquisa - precisa estar continuamente inserida nos fluxos de geração de conhecimento científico, estando presente em todo o ciclo de vida dos dados, que se inicia na concepção do projeto, fluindo por todo o ciclo de vida da pesquisa (YU; DEUBLE; MORGAN, 2017). A abordagem do ciclo de vida dos dados será apresentada na próxima seção.

2.5. CICLO DE VIDA DOS DADOS

Os materiais digitais como um todo, incluindo os dados de pesquisa, são suscetíveis às mudanças tecnológicas desde o momento de sua criação. As atividades, tanto de curadoria quanto de preservação, devem ser realizadas sob abordagem do ciclo de vida dos dados, garantindo que todos os estágios necessários sejam identificados e planejados e que as ações sejam implementadas na sequência correta (HIGGINS, 2008).

O ciclo de vida dos dados, também chamado de ciclo de vida das informações, descreve os processos e estados dos dados (ou informações) (KOWALCZYK, 2018). Busca “[...] garantir a manutenção da autenticidade, confiabilidade, integridade e usabilidade do material digital, o que, por sua vez, garante a maximização do investimento em sua criação” (HIGGINS, 2008, p. 135, tradução nossa).

A ideia de ciclo de vida dos dados foi promovida como conceito para apoiar práticas de preservação digital e de curadoria durante a década de 1990 e início da década de 2000, obtendo popularidade à medida que a cultura de compartilhamento de dados tem se tornado parte da linguagem de pesquisa cotidiana (CORTI *et al.*, 2014).

Hoje, as organizações adotam modelos de ciclo de vida dos dados “[...] buscando desenvolver sistemas e promover boas práticas de gerenciamento, organização e preservação de dados de pesquisa” (CARLSON, 2014, p. 63, tradução nossa).

2.5.1. Modelos de Ciclo de vida dos dados

Os modelos de ciclo de vida “[...] representam o curso de vida de um sistema maior, como o processo de pesquisa, por meio de uma série de etapas ou fases sequencialmente relacionadas nas quais a informação é produzida ou manipulada” (HUMPHREY, 2006, p. 1, tradução nossa). Os modelos de ciclo de vida tendem a ser de natureza progressiva e circular. Eles são usados para “[...] demonstrar as conexões entre os estágios individuais e o ciclo como um todo” (CARLSON, 2014, p. 65-66, tradução nossa).

Diferentes modelos de ciclo de vida de pesquisa¹⁰ (*research life cycle models*) são relevantes para o desenvolvimento de serviços de dados nas bibliotecas (CARLSON, 2014; HUMPHREY, 2006). “Ser capaz de visualizar serviços individuais e identificar como eles se encaixam em uma estrutura holística pode ajudar a identificar quaisquer lacunas em

¹⁰ Modelos projetados para ilustrar o processo geral de pesquisa, desde o seu início até o fim (CARLSON, 2014).

potencial ou abordar quaisquer áreas de fraqueza nos serviços” (CARLSON, 2014, p. 65-66, tradução nossa). O modelo é uma ferramenta visual que oferece perspectiva de alto nível das ações, operações ou processos individuais que devem ser realizados em diferentes etapas, ajudando a antecipar e planejar ações específicas a serem tomadas em cada estágio (FAUNDEEN *et al.*, 2013).

A pesquisa contém uma série de atividades inter-relacionadas. As atividades centrais ou para apoiar o processo de investigação merecem o desenvolvimento de modelos próprios, assim como os processos de aquisição e manutenção de financiamento para a pesquisa e os de atividades de geração e uso dos dados. Os modelos de ciclo de vida dos dados são subconjunto dos modelos de ciclo de vida de pesquisa, abrangendo todo o processo de pesquisa, destacando os aspectos ou atributos que pertencem aos dados (CARLSON, 2014), estendendo o ciclo de pesquisa típico (CORTI *et al.*, 2014).

De acordo com Pouchard (2015, p. 180, tradução nossa), “os modelos de ciclo de vida de dados apresentam estrutura para organizar as tarefas e atividades relacionadas à gestão de dados dentro de um projeto ou de uma organização”. Esses modelos sugerem um arranjo que permite considerar as diversas atividades que serão executadas durante todo o ciclo de vida dos dados. O modelo de ciclo de vida facilita as ações de curadoria, desde que preparadas com antecedência, mesmo antes da criação dos dados (BALL, 2012).

O uso de modelos de ciclo de vida para dados pode fornecer uma estrutura útil para articular os estágios pelos quais os dados passarão – transformações em seu formato, aplicação, uso e finalidade -, para contextualizar e comunicar tipos de serviços de dados que podem ser fornecidos, de modo a atender às situações e necessidades reais das comunidades (CARLSON, 2014; POUCHARD, 2015).

Carlson (2014) aponta benefícios da aplicação de modelos de ciclo de vida na gestão e curadoria de dados, uma vez que esses procedimentos são complexos. A aplicação dos modelos ajuda a definir e ilustrar visualmente os processos, facilitando a identificação das partes componentes ou etapas distintas dos dados, identificando as necessidades específicas do pesquisador. Além disso, “dividir o processo em estágios de componentes também ajuda a identificar e contabilizar os papéis, responsabilidades, marcos e outros componentes-chave na oferta de serviços de dados” (CARLSON, 2014, p. 65, tradução nossa).

Além de fornecer estrutura para identificação e desenvolvimento de serviços de dados, Carlson (2014) e Pouchard (2015) apontam que os modelos de ciclo de vida de dados ajudam a comunicar esses serviços aos usuários – pesquisadores, gerentes de dados,

curadores, especialistas em repositórios, bibliotecários e gerentes encarregados de organizar os dados em um projeto ou em um laboratório -, transmitindo-os de maneira atraente e inteligível, devido a sua natureza visual.

Eles [os modelos de ciclo de vida dos dados] servem como ponto de referência comum na negociação do que, quando e como os serviços serão entregues a um determinado grupo. Demonstrem as relações entre os serviços e os pré-requisitos ou outros requisitos necessários para que os consumidores aproveitem os serviços (CARLSON, 2014, p. 66, tradução nossa).

Os modelos de ciclo de vida de dados apresentam limitações que precisam ser reconhecidas. Tendem a projetar o ideal e não a realidade. Nesses modelos, o ciclo de vida dos dados é ordenado e linear. O processo, porém, pode não fluir tão bem, uma vez que os dados são produto da pesquisa - um empreendimento confuso que envolve lidar com o inesperado e responder a erros. Os modelos mascaram a complexidade do trabalhar com dados; representam vários projetos de pesquisa ou empreendimentos, ignorando a heterogeneidade e a diversidade de abordagens e práticas presentes nos projetos; geralmente refletem as perspectivas, interesses e vieses das organizações que os criaram (CARLSON, 2014).

Tal fato é fácil de entender, pois 'modelos' são paradigmas, 'tipos ideais' que orientam o processo, mas que não impede a ocorrência de fatos inesperados, inerentes a dinâmica de qualquer processo.

De acordo com Carlson (2014), três tipos de modelos de ciclo de vida são percebidos no desenvolvimento e suporte de serviços de dados, os modelos baseados em indivíduos, organizações ou comunidades:

- c) **modelos baseados em indivíduos** - compreendem um projeto de pesquisa específico, exibindo os seus detalhes e suas complexidades. Podem servir como planejamento eficaz ou ferramenta organizacional para projetar e executar um projeto de pesquisa (quais, quando e quem deve realizar as atividades necessárias). É preciso capturar informações sobre os dados gerados pelo projeto de pesquisa e identificar as demandas do pesquisador na gestão e organização desses dados;
- d) **modelos baseados em organizações** - semelhantes aos modelos baseados em indivíduos, contudo com propósitos diferentes. São produzidos por organizações que oferecem serviços ou assistência a pesquisadores (bibliotecas, repositórios de dados, sociedades acadêmicas, editores e outros). Os diferentes estágios do modelo do ciclo de vida geralmente se propõem a referenciar os serviços prestados pela

organização e instruir os pesquisadores sobre as etapas que devem ser tomadas na preparação para utilizá-los;

- e) **modelos baseados em comunidades** – servem como uma visão geral de alto nível, representando os componentes das melhores práticas e suas conexões. São desenvolvidos para apoiar ou atender às necessidades de uma comunidade de pesquisa específica. Tem como objetivo transmitir uma visão compartilhada das melhores práticas para o público interessado. É acompanhado por relatório ou outra publicação que fornece descrição mais detalhada das práticas.

As instituições optam por desenvolver seus próprios modelos para conceituarem e representarem os serviços de dados ou adotar modelo criado por outra instituição. Caso optem por desenvolver um modelo, deve ser considerado o escopo dos serviços de dados, estabelecendo com qual o nível de serviços e com que público-alvo (CARLSON, 2014).

Se o serviço de dados for focado em projetos, o modelo do ciclo de vida baseado no indivíduo é a opção mais apropriada. Ele fornece uma compreensão mais detalhada do trabalho executado. Os serviços devem abordar os problemas que podem surgir na gestão e na curadoria dos dados produzidos à medida que progridem no ciclo de vida (CARLSON, 2014).

Se o serviço de dados for destinado a atender um público mais amplo do que um projeto de pesquisa individual, deve-se escolher um modelo de ciclo de vida baseado na organização. Esse tipo de modelo representa fusão de ciclos de vida de dados baseados em indivíduos, fornecendo representação mais ampla de estágios ou atividades que são comuns em diferentes projetos de pesquisa. A precisão e o nível de detalhes irão depender do número e variedade de projetos de pesquisa, campos ou disciplinas representados (CARLSON, 2014).

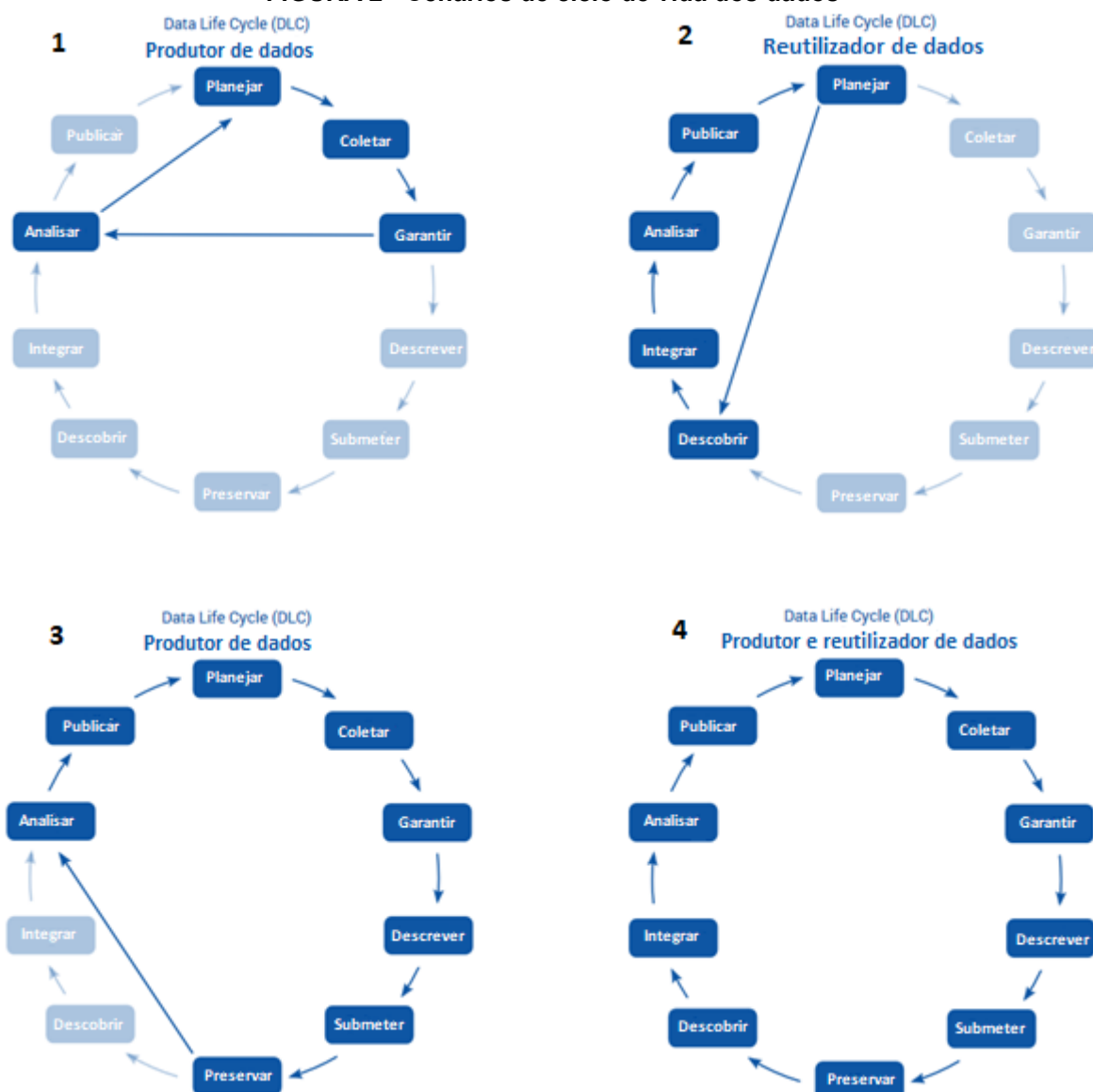
Qualquer que seja a escolha, o modelo de ciclo de vida deve incorporar as melhores práticas identificadas nos modelos de ciclo de vida baseados na comunidade. O grau de especificações que será adotado dependerá dos tipos de serviços (curadoria, preservação, publicação *etc.*) que a instituição pretende oferecer (CARLSON, 2014).

Caso opte por adotar um modelo de ciclo de vida gerado por outra instituição, deve-se ter a cautela de “[...] garantir que o modelo e suas partes componentes sejam adaptados e reflitam com precisão as práticas de pesquisa e dados dos grupos que pretendem representar” (CARLSON, 2014, p. 77, tradução nossa).

Orienta o *Digital Curation Center* (DCC) que, dependendo da demanda, as instituições podem iniciar em qualquer estágio do ciclo de vida (DCC, 2018a), uma vez que o uso do ciclo por completo “[...] pode variar dependendo do tipo de pesquisa conduzida, do tipo de dados coletados e do tipo de pesquisadores” (GFBIO, 2018, *online*, tradução nossa). Esse fato pode ser observado nos modelos de ciclo de vida dos dados apresentado pelo consórcio da *German Federation for Biological Data* (GFBio).

O consórcio GFBio apresenta quatro cenários de ciclo de vida dos dados com base no tipo de pesquisador, como produtor e como reutilizador de dados, orientando para a escolha daquele que melhor se encaixe a demanda. Os quatro cenários de ciclo de vida dos dados são apresentados na FIG. 2.

FIGURA 2 - Cenários de ciclo de vida dos dados



Fonte: Adaptado de GFBIO (2018, *online*).

Há vários modelos possíveis para descrever o ciclo de vida dos dados. Todos são divididos em etapas ou estágios sucessivos, sendo as principais: criação, captura (entrada em um sistema de informação ou banco de dados), armazenamento e manutenção, uso e descarte. Contudo, esse número pode variar (TJALSMA; ROMBOUTS, 2011).

Diferentes modelos de ciclos de vida apresentam diferentes níveis de granularidade e detalhes (KOWALCZYK, 2018). Corti *et al.* (2014) sintetizam as principais etapas e atividades relacionadas aos dados normalmente realizadas no ciclo de vida dos dados da pesquisa, conforme quadro 7.

QUADRO 7 - Atividades típicas realizadas no ciclo de vida dos dados de pesquisa

Etapas	Principais atividades
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
	Analisar
Publicação e compartilhamento	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
	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 pela autora com base em Corti *et al.* (2014, p. 17-18).

O *Data Management Working Group* (DMWG) da *United States Geological Survey* (USGS) investigou os modelos de ciclo de vida de dados existentes, resultando na compilação de mais de 50 modelos. Entre eles, os acadêmicos, os de órgãos federais e os de organizações internacionais. O DMWG apontou que “os modelos diferiram amplamente em granularidade, complexidade, apresentação e perspectiva da comunidade [...]” (FAUNDEEN *et al.*, 2013, p. 1, tradução nossa).

Para ilustrar a diversidade de modelos de ciclo de vida dos dados, apresentamos alguns exemplos como o *DCC Curation Lifecycle Model*, *DDI Combined Life Cycle Model*, *DataONE Data Life Cycle*, *USGS Science Data Lifecycle Model* e *Research Data Lifecycle – UK Data Service*. Embora contenham variações, os modelos evidenciam as principais etapas e atividades identificadas por Corti *et al.* (2014).

2.5.1.1. DCC Curation Lifecycle Model

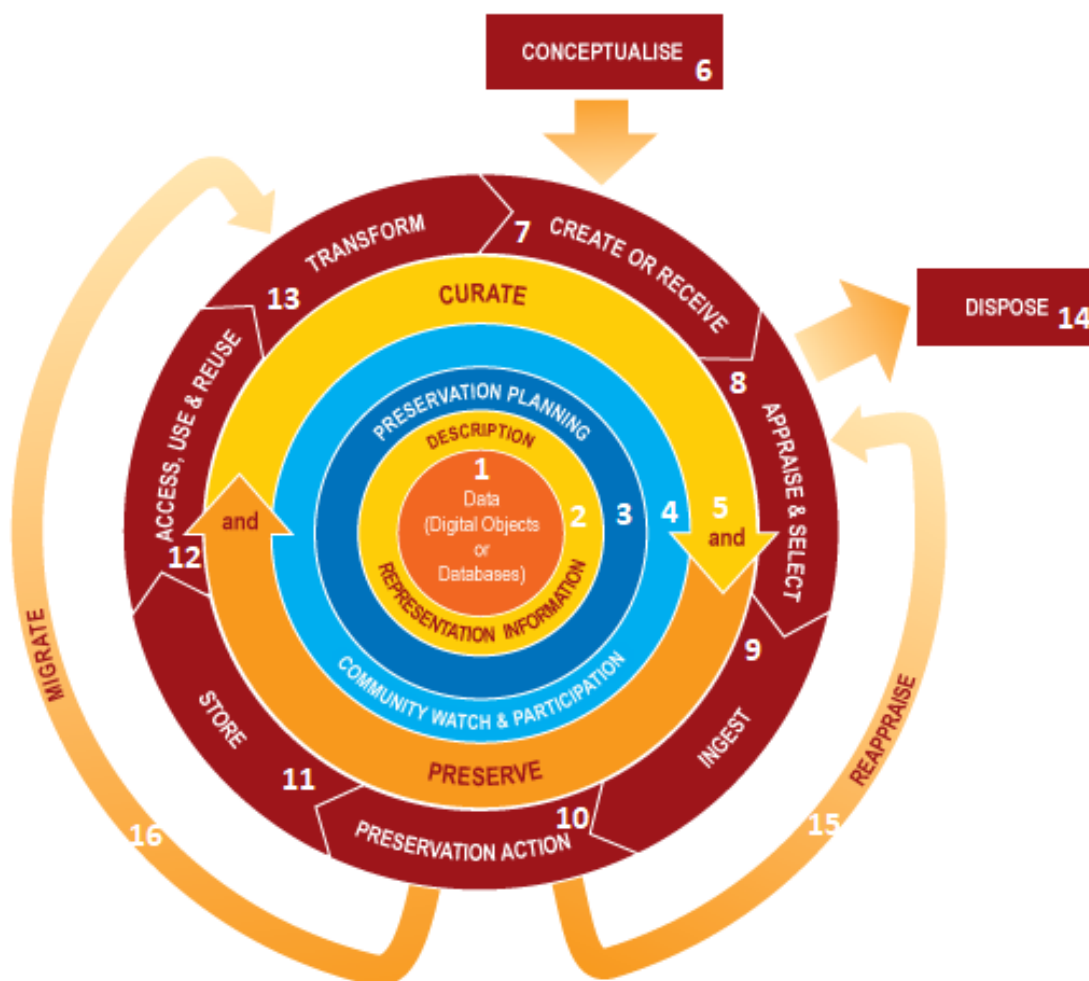
O modelo de ciclo de vida de curadoria do *Digital Curation Centre* (DCC) (FIG. 3) “[...] fornece uma visão geral gráfica de alto nível dos estágios exigidos para a curadoria bem-sucedida e a preservação de dados desde a conceituação inicial ou recebimento até o ciclo de curadoria iterativa” (DCC, 2018b, *online*, tradução nossa).

Para Higgins (2008, p. 135, tradução nossa), o *DCC Curation Lifecycle Model* é modelo genérico, indicativo e não exaustivo, portanto “[...] adaptável a diferentes domínios e extensível para permitir que atividades de curadoria e preservação sejam planejadas em diferentes níveis de granularidade”. Dessa maneira, o modelo pode ser empregado para definir os “[...] papéis e responsabilidades; construir estruturas de padrões e tecnologias; e garantir que os processos e políticas sejam adequadamente documentados”.

O modelo pode ser usado para planejar atividades dentro de um projeto de pesquisa, organização ou consórcio específico garantindo que todas as etapas necessárias sejam realizadas, cada uma na sequência correta (DIGITAL CURATION CENTRE, 2018b).

O DCC orienta que esse modelo pode ser usado para planejar atividades dentro da instituição, garantindo que todas as etapas necessárias no ciclo de vida da curadoria sejam incluídas. Contudo, as instituições usuárias do modelo podem entrar em qualquer estágio do ciclo de vida, dependendo de sua demanda (DCC, 2018a).

FIGURA 3 - DCC Curation Lifecycle Model



Fonte: Adaptado de DCC (2018a).

Nota:

- 1- Dados (Objetos digitais ou base de dados)
- 2 - Descrição e representação da informação
- 3 – Planejamento da preservação
- 4 – Observação e participação
- 5 – Curadoria e preservação
- 6 – Conceitualizar
- 7 – Criar e receber
- 8 – Avaliar e selecionar
- 9 – Ingerir
- 10 – Ações de preservação
- 11 – Armazenar
- 12 – Acessar, usar e reusar
- 13 – Transformar
- 14 – Descartar
- 15 – Reavaliar
- 16 - Migrar

De acordo com Higgins (2008), o modelo proposto pelo DCC identifica as ações de curadoria que são aplicáveis em todo o ciclo de vida digital, as ações que precisam ser realizadas sequencialmente e aquelas que são realizados ocasionalmente, dependendo da circunstância.

O centro do modelo é composto pelos dados, qualquer informação em formato digital - objetos digitais (ex.: arquivos textuais, imagens ou arquivos de som, juntamente com identificadores e metadados; *websites* e outros) ou bancos de dados estruturados (HIGGINS, 2008).

Além dos dados, o modelo é composto por três tipos de ações (DCC, 2018c):

1. **ações do ciclo de vida completo (*Full Lifecycle Actions*)** – atividades que acontecem a qualquer momento durante o ciclo de vida da curadoria digital. São relevantes para muitas **ações sequenciais** diferentes. São mostradas em anéis concêntricos ao redor dos objetos de dados no centro do modelo (Quadro 8).

QUADRO 8 - Ações do ciclo de vida completo

Ações do ciclo de vida completo	
Descrição e representação da informação	Atribuir metadados administrativos, descritivos, técnicos, estruturais e de preservação, usando padrões apropriados, para assegurar uma descrição e controle adequados em longo prazo. Coletar e atribuir informações de representação necessárias para entender e preparar o material digital e os metadados associados.
Planejamento da preservação	Planejar a preservação durante todo o ciclo de vida da curadoria de material digital. Isso incluiria planos para gestão e administração de todas as ações de ciclo de vida de curadoria.
Observação e participação	Manter vigilância sobre atividades comunitárias apropriadas e participar do desenvolvimento de padrões compartilhados, ferramentas e <i>software</i> adequado.
Curadoria e preservação	Estar ciente e realizar ações de gestão e administrativas, planejadas para promover a curadoria e a preservação durante todo o ciclo de vida da curadoria.

Fonte: (HIGGINS, 2008, p. 137, tradução nossa).

2. **ações sequenciais (*Sequential Actions*)** – sequência que forma a base da cadeia de curadoria, sendo executada enquanto os dados estiverem sendo curados. A sequência é composta por etapas que são repetidamente realizadas para garantir que os dados sejam selecionados de acordo com as melhores práticas (Quadro 9).

QUADRO 9 - Ações sequenciais

Ações sequenciais	
Conceitualizar	Conceber e planejar a criação de dados, incluindo método de captura e opções de armazenamento.
Criar e 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 políticas de coleta documentadas, de criadores de dados, outros arquivos, repositórios ou centros de dados e, se necessário, atribuir metadados apropriados.
Avaliar e selecionar	Avaliar os dados e selecionar para curadoria e preservação a longo prazo. Seguir as orientações documentadas, políticas ou requisitos legais.
Ingerir	Transferir dados para um arquivo, repositório, <i>data center</i> ou outro custodiante. Seguir as orientações documentadas, políticas ou requisitos legais.

Ações de preservação	Realizar ações para garantir a preservação a longo prazo e a retenção da natureza autoritativa dos dados. As ações de preservação devem garantir que os dados permaneçam autênticos, confiáveis e usáveis, 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, seguindo os padrões relevantes.
Acessar, usar e reusar	Certificar-se de que os dados estejam acessíveis para os usuários designados e para os reutilizadores, no dia-a-dia. Isso pode ser na forma de informações tornadas publicamente disponíveis. Controles de acesso robustos e procedimentos de autenticação podem ser aplicáveis.
Transformar	Criar novos dados do original, por exemplo.
	- Por migração para um formato diferente.
	- Criando um subconjunto, por seleção ou consulta, para criar resultados recém-derivados, talvez para publicação.

Fonte: (HIGGINS, 2008, p. 138, tradução nossa).

3. **ações ocasionais (*Occasional Actions*)** – ações que interrompem ou reordenam as ações sequenciais como resultado de uma decisão (Quadro 10).

QUADRO 10 - Ações ocasionais

Ações ocasionais	
Descartar	Descartar os dados, que não foram selecionados para curadoria e preservação de longo prazo, de acordo com políticas documentadas, orientação ou requisitos legais. Normalmente, os dados podem ser transferidos para outro arquivo, repositório, <i>data center</i> ou outro custodiante. Em alguns casos, os dados são destruídos. A natureza dos dados pode, por razões legais, necessitar de destruição segura.
Reavaliar	Retornar dados que falham em procedimentos de validação para avaliação adicional e nova seleção.
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 contra a obsolescência de <i>hardware</i> ou <i>software</i> .

Fonte: (HIGGINS, 2008, p. 138, tradução nossa).

2.5.1.2. DDI Combined Life Cycle Model

O modelo de ciclo de vida combinado da *Data Documentation Initiative* (DDI, 2018) (FIG. 4) é voltado, particularmente, para dados das áreas de ciências sociais. Possui oito elementos distribuídos quase que totalmente de forma linear, com apenas um caminho alternativo e um *loop* de *feedback*.

FIGURA 4 - DDI Combined Life Cycle Model
DDI Data Lifecycle



Fonte: DDI (2018, *online*, tradução nossa).

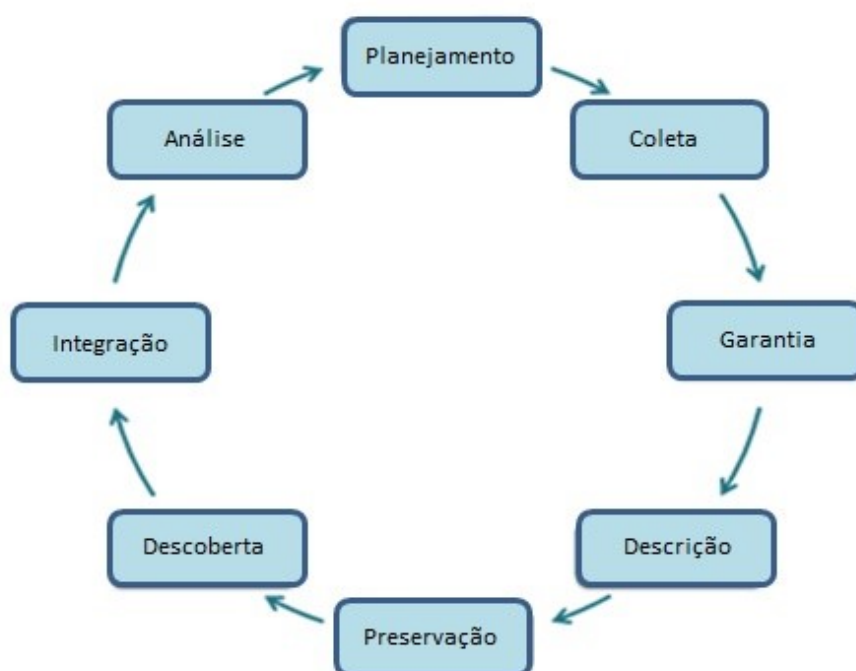
Os oito elementos são apresentados a seguir (BALL, 2012):

1. conceito de estudo – estágio de planejamento da pesquisa. Inclui não apenas a escolha da pergunta de pesquisa e a metodologia para coletar os dados, mas também a maneira pela qual os dados serão processados, analisados e utilizados para responder à pergunta;
2. coleta de dados (métodos e fontes de coleta) - realizada por meio de pesquisas, censos, votação ou registros de saúde, estatísticas de comércio ou coleções baseadas na *web*;
3. processamento de dados – os dados de entrada são processados e analisados para produzir dados de saída que respondam à questão de pesquisa;
4. arquivamento de dados (para garantir acesso de longo prazo) – os dados devem ser passados para um arquivo, em vez de apenas mantidos por pesquisadores. O arquivo não apenas preserva os dados (e metadados) como também agrega valor a eles ao longo do tempo;
5. distribuição de dados – distribuição direta aos usuários ou por meio de uma biblioteca ou arquivo de dados;
6. descoberta de dados - divulgados através de livros, publicações de periódicos, páginas ou outros serviços *online*;
7. análise de dados - os dados podem ser usados por outros dentro dos limites da conceituação original;
8. reaproveitamento - os dados podem ser usados dentro de uma estrutura conceitual diferente; exemplos incluem amostragem ou reestruturação dos dados, combinando os dados com outros conjuntos semelhantes ou produzir materiais pedagógicos.

2.5.1.3. DataONE Data Life Cycle

O ciclo de vida dos dados DataONE (FIG. 5) foi desenvolvido pela *Data Observation Network for Earth* (DataONE) e “[...] fornece uma visão geral de alto nível dos estágios envolvidos no gerenciamento bem-sucedido e na preservação de dados para uso e reutilização”(DataOne, 2018, *online*).

FIGURA 5- DataONE Data Life Cycle



Fonte: (DataONE, 2018, *online*, tradução nossa).

O ciclo de vida dos dados DataONE compõe-se de oito etapas (BALL, 2012; DataOne, 2018) que visam:

1. planejamento – criar um plano de gestão de dados para controlar como os dados são tratados em todo o projeto de pesquisa. Descrever como os dados serão compilados, gerenciados e disponibilizados ao longo de sua vida útil;
2. coleta – adquirir dados por meio de observações feitas à mão, sensores ou outros instrumentos de laboratório e medição. Os dados devem ser organizados em planilhas ou bancos de dados;
3. garantia de qualidade - realizar verificações e inspeções para garantir a excelência dos dados;
4. descrição: os dados são descritos com precisão e detalhadamente usando os padrões de metadados apropriados;

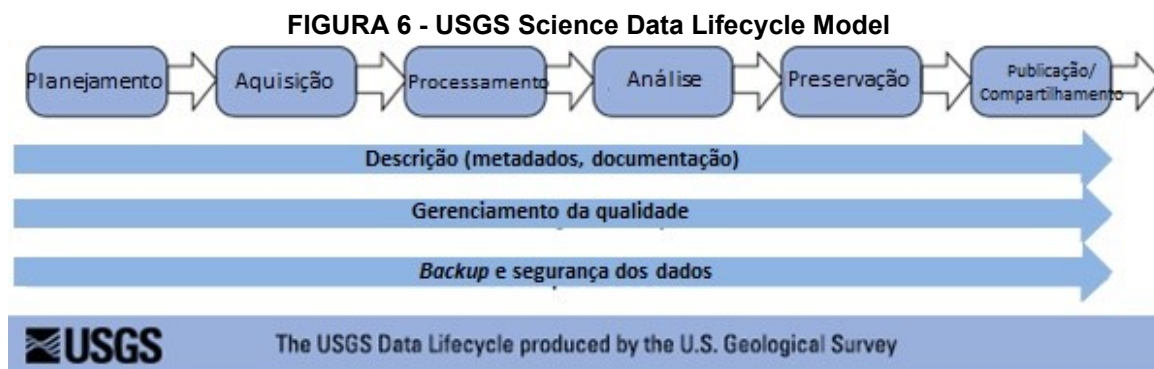
5. preservação – depositar os dados em um local apropriado (*data center* ou repositório) para que eles possam ser protegidos de degradação e obsolescência de formato;
6. descoberta – expor os metadados de descoberta em uma interface pesquisável de modo a localizar dados potencialmente úteis;
7. integração - combinar dados de diferentes fontes para formar um conjunto homogêneo de dados que possam ser prontamente analisados. Dar uma representação comum para um conjunto de dados diferentes (formato, esquema de codificação, ontologia), preservando a proveniência;
8. análise - aplicar modelos estatísticos e analíticos aos dados de forma a extrair respostas significativas para questões de pesquisa.

De acordo com o DataOne (2018), algumas atividades podem usar apenas parte do ciclo de vida (uma metanálise pode se concentrar nas etapas 6 a 8, enquanto um projeto focado na análise de dados primários pode ignorar as etapas 6 e 7) ou, ainda, podem não seguir o caminho de forma linear.

2.5.1.4. USGS Science Data Lifecycle Model

Desenvolvido pelo *United States Geological Survey* (USGS), o *USGS Science Data Lifecycle Model* (FIG. 6) ajuda os cientistas “[...] a entenderem e gerenciarem o ciclo de vida dos dados e produtos de informação para garantir que os dados sejam detectáveis, bem descritos e preservados para acesso e uso além da vida de projetos de pesquisa”. (FAUNDEEN *et al.*, 2013, p. 2, tradução nossa).

O modelo é formado por seis elementos principais – abordam atividades e saídas distintas exclusivas do estágio (planejar, adquirir, processar, analisar, preservar e publicar/compartilhar) e três elementos transversais - atividades críticas que devem ser realizadas continuamente em todos os estágios do ciclo de vida para ajudar no suporte à gestão de dados efetivo (descrever – metadados e documentação-, gerenciar a qualidade e *backup* e segurança), conforme FIG. 6 (FAUNDEEN *et al.*, 2013).



Fonte: (USGS, 2018b, tradução nossa).

Elementos principais (FAUNDEEN *et al.*, 2013; USGS, 2018b):

1. planejamento – leva em consideração todas as atividades relacionadas ao manuseio dos ativos de dados, desde o início do projeto até a publicação e o arquivamento, bem como os recursos necessários, incluindo financiamento, recursos técnicos e de pessoal. Na saída, há uma sequência documentada de ações planejadas para identificar e proteger recursos, coletar, manter, proteger e utilizar os dados armazenados, na forma de plano de gestão de dados;
2. aquisição – envolve as atividades pelas quais dados novos ou existentes são coletados, gerados, ou considerados e avaliados para reutilização. A aquisição de dados ocorre de quatro formas: coleta de novos dados; conversão/transformação de dados legados; compartilhamento/troca dados; e compra de dados. As saídas desse elemento são as entradas de dados do projeto;
3. processamento – abrange um conjunto de atividades estruturadas, resultando na alteração ou integração de dados. Engloba as ações ou etapas executadas para verificar, organizar e validar os dados, transformar, subdividir, resumir, integrar, derivar, entre outras. Os métodos de processamento devem ser rigorosamente documentados para garantir a utilidade e integridade dos dados. As saídas desse elemento são conjuntos de dados prontos para integração e análise;
4. análise – envolve as atividades associadas à exploração e interpretação de dados processados, descrição de fatos, detecção de padrões, desenvolvimento de explicações, teste de hipóteses, gerando descobertas e conclusões. As atividades analíticas incluem resumo, gráficos, análise estatística, análise espacial e modelagem, visualização, análise de imagem, interpretação, entre outras, e são usadas para produzir resultados e informações científicas. As saídas desta fase são interpretações ou novos conjuntos de dados, publicados em relatórios escritos, ou, em formatos legíveis por máquina, como camadas de mapa ou resultados de modelagem numérica;

5. preservação – representa as ações e procedimentos associadas ao armazenamento de dados para uso e acessibilidade em longo prazo, incluindo o arquivamento e/ou envio de dados para repositório de dados. Deve-se planejar a preservação a longo prazo de dados, metadados, produtos auxiliares, formatos de armazenamento e qualquer documentação adicional, para garantir disponibilidade e reutilização;
6. publicação/compartilhamento – os dados, assim como as publicações tradicionais, são produtos de pesquisa e devem ser publicados/compartilhados, sendo esse elemento parte importante do ciclo de vida. A publicação/compartilhamento combina os conceitos da publicação tradicional revisada por especialistas com a distribuição de dados por meio de *sítes*, catálogos de dados, mídias sociais e outros locais. Contudo, devem ser garantidos controles para proteção dos dados proprietários e pré-decisórios e a integridade dos dados em si. Requer a disponibilização de metadados completos para que os dados sejam úteis àqueles que os estão recebendo.

Elementos transversais (FAUNDEEN *et al.*, 2013; USGS, 2018b):

1. descrição (metadados, documentação) – evidencia a importância da documentação durante todo o ciclo de vida dos dados, enfatizando o passo a passo em cada estágio do ciclo de vida, desde o plano de gestão de dados. A documentação deve possuir detalhes suficientes para que outros cientistas possam validar os resultados da pesquisa por meio de replicação, avaliar a validade dos resultados e determinar a utilidade dos dados para pesquisas futuras. A documentação deve ser atualizada de modo a refletir as ações tomadas durante todo o ciclo. Além disso, a inclusão de metadados e documentação baseada em padrões facilitam a indexação, adesão, compreensão e usos futuros dos dados. A atualização e completude dos metadados são essenciais para manter a qualidade dos dados;
2. gerenciamento da qualidade – uso de protocolos para garantir que os dados sejam coletados, manipulados, processados, usados e mantidos adequadamente em todos os estágios do ciclo de vida dos dados científicos, garantindo sua qualidade. Esse elemento evoca a necessidade de os cientistas planejarem medidas que garantam a qualidade dos dados desde o início do projeto, monitorando e ajustes contínuos do controle de qualidade nas etapas subsequentes do ciclo de vida para verificar se essas medidas funcionam conforme o esperado;
3. segurança dos dados – utilização de medidas para prevenção de perdas físicas de dados - perda acidental, falhas de *hardware* ou *software*, desastres naturais, erros

humanos, corrupção e acesso não autorizado. Tais medidas incluem *backups* rotineiros dos dados, garantindo que eles estejam acessíveis. Esse elemento lembra os pesquisadores de planejar serviços seguros para o compartilhamento de dados.

2.5.1.5. Research Data Lifecycle – UK Data Service

Modelo desenvolvido pelo *UK Data Service* (FIG. 7), serviço criado pelo *Economic and Social Research Council* (ESRC) do Reino Unido. Esse é um modelo mais genérico (KOWALCZYK, 2017), possuindo seis etapas, compostas pelas seguintes atividades (UK DATA SERVICE, 2018):

FIGURA 7 - Research Data Lifecycle: UK Data Service



Fonte: (UK DATA SERVICE, 2018, *online*, tradução nossa).

1. planejamento da pesquisa – projetar a pesquisa; planejar a gestão de dados; planejar o consentimento para compartilhamento de dados; planejar a coleta, os protocolos e modelos de processamento e; explorar as fontes de dados existentes;
2. coleta de dados – coletar os dados; capturar dados com metadados; adquirir dados de terceiros;

3. processamento e análise de dados - digitar, digitalizar, transcrever e traduzir dados; verificar, validar, limpar, tornar anônimo; derivar dados; descrever e documentar dados; gerenciar e armazenar; analisar e interpretar; produzir *outputs* de pesquisa; e citar fontes de dados;
4. publicação e compartilhando de dados – estabelecer direitos autorais; criar documentação do usuário; criar metadados de descoberta; selecionar a forma de acesso apropriado aos dados; publicar/compartilhar dados; e promover dados;
5. preservação dos dados – migrar dados para o melhor formato/mídia; armazenar e fazer *backup*; criar documentação de preservação; e preservar e curar dados;
6. reutilização dos dados - realizar análises secundárias; realizar pesquisas de acompanhamento; realizar análises de pesquisa; examinar as descobertas; e usar dados para ensinar e aprender.

2.6. COMPARTILHAMENTO E REUSO DE DADOS

O compartilhamento de dados “[...] é a prática de disponibilizar dados usados para pesquisa acadêmica para outros pesquisadores” (NNLM, 2018, *online*, tradução nossa).

O compartilhamento de dados e as políticas que o exigem ou incentivam tomam como base o pressuposto de que os dados são úteis para os outros (BORGMAN, 2012). A prática ou necessidade de compartilhamento de dados não é nova, conforme já apontado pelo relatório do *Committee on National Statistics* (CNSTAT) (FIENBERG; MARTIN; STRAF, 1985). Contudo, ganhou ênfase com o surgimento da ciência intensiva em dados e continua a crescer a medida que o movimento do acesso aberto ganha força em todo o mundo (KIM, 2013).

É esperado que os cientistas disponibilizem seus dados, pois a maioria das pesquisas são financiadas diretamente pelos governos ou, indiretamente, por meio dos salários dos pesquisadores, por fundos públicos ou, ainda, os dados são cobrados em favor do bem público comum como saúde pública, monitoramento ambiental e segurança (COSTELLO, 2009). O compartilhamento de dados deve ser realizado, pois a transparência e disponibilidade da pesquisa são consideradas uma parte importante da descoberta científica (NNLM, 2018).

O compartilhamento de dados pode acontecer de maneira informal ou formal, conforme apresentado por Clubb *et al.* (1985). O compartilhamento informal acontece quando um pesquisador ou uma instituição solicita e recebe cópias de dados de outros pesquisadores

ou organizações. Os dados podem ser fornecidos a outros pesquisadores. O compartilhamento formal acontece em cenário mais estruturado (CURTY, 2015), envolvendo o uso de “[...] instalações intermediárias que funcionam como repositórios de dados e serviços de disseminação” (CLUBB *et al.*, 1985, p. 63, tradução nossa).

De acordo com Gore (2011), idealmente, os dados podem ser compartilhados com outros cientistas para fins de reutilização e reanálise, de forma a acelerar o processo de descoberta científica. O compartilhamento de dados traz benefícios como apontados no Quadro 11:

QUADRO 11- Benefícios do compartilhamento de dados

Benefícios	AUTOR (ano)
Acelerar o processo de descoberta científica	(GORE, 2011)
Alavancar investimentos	(POOLE, 2016)
Aumentar a citação	(POOLE, 2016)
Contribuir para o desenvolvimento de conhecimento teórico e conhecimento de técnica analítica	(FIENBERG; MARTIN; STRAF, 1985)
Diminuir a necessidade de repetição de trabalho	(LEWIS, 2010)
Disponibilizar ao público os resultados da pesquisa financiada publicamente	(BORGMAN, 2012; POOLE, 2016)
Evitar custos de coleta de dados	(CALLAGHAN <i>et al.</i> , 2012; FIENBERG; MARTIN; STRAF, 1985)
Fornecer recursos para treinamento em pesquisa e melhorar a aprendizagem (aspecto pedagógico)	(FIENBERG; MARTIN; STRAF, 1985; POOLE, 2016)
Fornecer um contexto descritivo ou histórico	(POOLE, 2016)
Garantir que a pesquisa não será perdida ou inadvertidamente destruída	(LEWIS, 2010)
Gerar novos conhecimentos, novos temas de pesquisa a partir da reanálise dos dados existentes ou comparações com novos dados e incentivar as múltiplas perspectivas	(BORGMAN, 2012; FIENBERG; MARTIN; STRAF, 1985; GORE, 2011; LEWIS, 2010; POOLE, 2016)
Incentivar o uso mais apropriado de dados empíricos na formulação e avaliação de políticas	(FIENBERG; MARTIN; STRAF, 1985)
Permitir a descoberta e reutilização de dados já coletados	(CALLAGHAN <i>et al.</i> , 2012)
Permitir novos caminhos metodológicos ou melhorar os métodos de medição e coleta de dados	(FIENBERG; MARTIN; STRAF, 1985; POOLE, 2016)
Permitir reproduzir, verificar, repetir, refutar, detectar fraudes, refinar os resultados originais e apurar a integridade da pesquisa	(BORGMAN, 2012; FIENBERG; MARTIN; STRAF, 1985; LEWIS, 2010; POOLE, 2016)
Permitir um avanço mais eficiente da ciência, pesquisa e inovação	(BORGMAN, 2012; FIENBERG; MARTIN; STRAF, 1985)
Possibilitar a meta-análise	(POOLE, 2016)
Promover o desenvolvimento econômico	(POOLE, 2016)
Proporcionar oportunidades de colaboração	(CALLAGHAN <i>et al.</i> , 2012)
Proteger contra dados defeituosos (fraudulentos, distorcidos, inadequados), evitando desperdício de recursos	(POOLE, 2016)
Reforçar a pesquisa científica aberta	(FIENBERG; MARTIN; STRAF, 1985; LEWIS, 2010)

Fonte: Elaborado pela autora.

O compartilhamento de dados representa o sucesso de todo o processo. O reuso de dados só pode acontecer depois de os meios de compartilhamento estarem implantados (CRAGIN *et al.*, 2010). Contudo, o compartilhamento parece ainda não estar disponível para ampla utilização prática (KIM, 2013). As práticas de compartilhamento de dados de pesquisa é algo

comum em apenas algumas áreas como Astronomia e Genômica. Em outros campos, elas variam amplamente (BORGMAN, 2012).

Cada disciplina apresenta suas próprias capacidades e práticas de compartilhamento de dados. De acordo com Kim (2013), diversos estudos mostraram a existência de amplas variações na cultura de compartilhamento. Cada disciplina possui diferentes estratégias e mudanças em suas práticas dessa natureza. “Além disso, a prática de compartilhamento de cada disciplina difere em termos do volume de dados produzidos, valor percebido de compartilhamento, restrições éticas, e assim por diante” (KIM, 2013, p. 497-498, tradução nossa).

Na *big science*, por exemplo, as áreas tendem a ter práticas de dados padronizadas. Em áreas da *small science*, o compartilhamento de dados não é comum ou esperado. Funciona como uma indústria artesanal onde os “[...] dados são trocados com base em relações profissionais e comunicação pessoal” (CRAGIN *et al.*, 2010, p. 4023, tradução nossa). Percebemos que, na *big science*, são mais comuns as práticas formais. Nas áreas da *small science* prevalecem as práticas informais.

O propósito do compartilhamento é a reutilização de dados. Cabe-nos, pois, apresentar uma definição para reuso de dados científicos. O *Data Thesaurus* aponta que “a reutilização de dados é conceito que envolve o uso de dados de pesquisa para uma atividade de pesquisa ou finalidade diferente daquela para a qual foi originalmente planejada” (NNLM, 2018, *online*, tradução nossa).

Curty (2016) descreve o reuso dos dados científicos

como toda nova aplicação de dados por meio de re-análise [*sic*] e replicação, ou como a combinação de diferentes conjuntos de dados por meio de integração ou meta-análise, a partir de novas perguntas de pesquisa, novos métodos de análise, com propósitos similares ou distintos daqueles empreendidos no estudo original, com ou sem a participação do reutilizador (CURTY, 2016, p. 3-4).

O *Data Thesaurus* aponta que a reutilização de dados pode ser promovida por meio do envio de dados para um repositório apropriado e pela garantia de que os dados estão bem descritos e compreensíveis para qualquer outro pesquisador que queira usá-los (NNLM, 2018).

Os dados de pesquisa por si “[...] são incompreensíveis e, portanto, inúteis, a menos que haja uma descrição detalhada e clara de como e quando foram coletados e como os dados derivados foram produzidos” (GRAY *et al.*, 2002, p. 106, tradução nossa). Contudo, para que sejam úteis, ou seja, reutilizados e atendam às diferentes demandas de pesquisa, os

dados devem seguir uma estrutura e organização claras e constituir coleções informativas relacionais e registradas em formato adequado ao tema tratado (CORRÊA, 2017).

Para que os usuários obtenham informações que permitam reutilizá-los, os dados científicos devem ser acompanhados de metadados que indiquem como os instrumentos foram projetados e construídos, quando, onde e como foram coletados, e uma descrição cuidadosa das etapas de processamento que levaram a criação de dados derivados (GRAY *et al.*, 2002).

A despeito dos benefícios apresentados pelo compartilhamento de dados (BORGMAN, 2012; CALLAGHAN *et al.*, 2012; FIENBERG; MARTIN; STRAF, 1985; GORE, 2011; LEWIS, 2010; POOLE, 2016), das tecnologias disponíveis, políticas, mandatos das agências financiadoras, movimento do acesso aberto e dados abertos – ciência aberta -, as práticas de compartilhamento progridem lentamente. Um dos principais motivos é que, geralmente, o compartilhamento acontece por meio de redes informais. Além disso, há outros motivos como: cautela em colocar dados no domínio público, questões relacionadas a formatos, padrões, metadados e políticas e direitos autorais (MACDONALD; MARTINEZ-URIBE, 2013). Outro problema seria a não recompensa percebida pelos pesquisadores. Nos últimos anos, tendência caminha do compartilhamento para a publicação de dados científicos (KRATZ; STRASSER, 2014).

2.7. PUBLICAÇÃO DE DADOS

Embora a comunidade científica, governos e agências de fomento advoguem e apreciem os benefícios da disponibilização de dados, as práticas de compartilhamento muitas vezes não são recompensadas (COSTELLO, 2009; KRATZ; STRASSER, 2015). Contudo, autores e administradores de dados buscam o reconhecimento pelo esforço intelectual que investem na criação de um bom conjunto de dados (PARSONS; FOX, 2013).

Diferentemente da publicação em mídia física, que se tornou sinônimo de processos de garantia da qualidade e longevidade do item publicado, na *internet*, pode-se publicar qualquer coisa. A publicação pode ser somente a disponibilização para *download*, uma forma menos confiável que a tradicional (LAWRENCE *et al.*, 2011).

No entanto, Callaghan *et al.* (2012) fazem uma distinção da publicação (grafada com p minúsculo) e Publicação (grafada com P maiúsculo). A publicação (p minúsculo) é vista como um processo relativamente fácil, onde os pesquisadores apenas disponibilizam os arquivos em um *site*, tornando-os abertos, sem qualquer forma de compromisso em longo prazo. É apenas a disponibilização para *download*, conforme apresentado por Lawrence *et al.* (2011).

Já a Publicação (P maiúsculo) é processo formal que agrega valor a um conjunto de dados, com compromisso com a persistência desses. Além disso, fornece mecanismo para garantir que o pesquisador obtenha crédito acadêmico por seu trabalho (CALLAGHAN *et al.*, 2012).

Neste trabalho, adotamos a segunda definição proposta por Callaghan *et al.* (2012) – publicação que agrega valor aos dados, com garantia de persistência - mesmo que grafada com letra minúscula.

De forma geral, todo processo de publicação deve executar cinco funções básicas (LAWRENCE *et al.*, 2011; VAN DE SOMPEL *et al.*, 2004):

1. registro;
2. certificação;
3. conscientização;
4. arquivamento;
5. recompensa.

De acordo com Van de Sompel *et al.* (2004), essas funções devem ser vistas numa perspectiva de cadeia de valor. O registro considera a data em que o manuscrito – ou conjunto de dados - foi recebido. O processo de revisão por pares certifica. Se o artigo – ou conjunto de dados - for publicado, cumpre a função de conscientização. A recompensa vem por meio das citações. O arquivamento é realizado de forma *ad hoc*, quando publicado por uma revista.

Além dessas funções básicas, em se tratando de publicação de dados, outra função deve ser executada (LAWRENCE *et al.*, 2011) a

6. definição do que está sendo publicado.

Essas seis funções – definição, registro, certificação, conscientização, arquivamento e recompensa - podem ser agrupadas em duas categorias (LAWRENCE *et al.*, 2011):

- a) ajudas à reutilização - tornam as publicações permanentemente disponíveis e o conhecimento dentro delas utilizáveis em outros contextos (Arquivamento, Conscientização e Definição).
- b) ativadores de reconhecimento - permitem medir e reconhecer o valor do trabalho (Registro, Recompensa e Certificação).

A lista de funções associadas à publicação é omissa em relação à propriedade dos dados e direitos autorais. Com os documentos, a lei de direitos autorais inclui o problema de

propriedade. Contudo, no caso dos dados, a questão do direito autoral e licenciamento não é clara, sendo usada uma variedade de licenças¹¹ pelos repositórios de dados (LAWRENCE *et al.*, 2011).

Os pesquisadores buscam nas publicações acadêmicas por serem validadas por processo científico bem estabelecido, um possível análogo para compartilhar e preservar dados (PARSONS; FOX, 2013).

A publicação de dados “[...] procura ser análoga à publicação de literatura acadêmica [...]” (PARSONS; FOX, 2013, p. 35, tradução nossa). Trabalha com o princípio, segundo o qual, os “[...] conjuntos de dados devem se juntar ao registro acadêmico e ter o mesmo *status* de primeira classe que os produtos de pesquisa tradicionais, como os artigos de periódicos” (KRATZ; STRASSER, 2014, p. 3, tradução nossa).

O objetivo da publicação de dados “[...] é ter conjuntos de dados como um ‘resultado de pesquisa de primeira classe’ que estará disponível, revisado por pares, citado, facilmente descoberto e reutilizável” (CANDELA *et al.*, 2015, p. 1748, tradução nossa), garantindo a reprodutibilidade da pesquisa e reutilização dos dados científicos.

O processo de publicação de dados apresenta como atores: autores, periódicos, arquivos e terceiros (consumidores de dados, citantes ou revisores) (LAWRENCE *et al.*, 2011).

Esses atores identificados por Lawrence *et al.* (2011) são chamados por Gray *et al.* (2002) de papéis no processo de publicação de dados, conforme apresentado a seguir:

- a) autores – produzem ou reúnem os dados, analisam e enviam os resultados baseados inteiramente em seus próprios experimentos para um periódico (*Publisher*);
- b) periódicos (*Publisher*) – recebem os resultados da pesquisa em forma de artigos. Documentam os dados de origem para que outros possam usá-los, bem como as etapas de processamento para que outros possam reproduzi-los;
- c) curadores – bibliotecas que armazenam e indexam as revistas;
- d) consumidores – outros cientistas que leem os artigos publicados nos periódicos e que podem reutilizar os dados.

¹¹ Creative commons (<https://creativecommons.org/licenses/>)
 Open Database License (ODbL) (<https://opendatacommons.org/licenses/odbl/>)
 Open Data Commons Attribution License (ODC-By) (<https://opendatacommons.org/licenses/by/1-0/index.html>)

A publicação de dados perpassa questões de preservação e curadoria. Uma vez publicados, os dados científicos devem permanecer disponíveis para sempre, para que outros cientistas possam reproduzir os resultados e fazer nova ciência com os dados. Contudo o processo é oneroso devido à (GRAY *et al.*, 2002):

- a) falta de padrões – poucas diretrizes para publicar dados, falta de padrões de metadados. O editor deve selecionar ou inventar seu próprio conjunto de metadados;
- b) processo laborioso – a documentação é pré-requisito para publicar dados. Documentar dados é trabalhoso e o sistema de recompensa não reconhece seu valor.

Lawrence *et al.* (2011) corrobora Gray *et al.* (2002) ao destacar a importância da permanência dos dados, que perpassa por três questões:

- a) a fragilidade das URLs¹², tornando impossível encontrar o material novamente na hipótese de ocorrer um problema com o identificador. Os autores sugerem o uso do DOI¹³ para fornecer identificador e localizador permanente para um conjunto de dados;
- b) garantia de que o conjunto de dados que foi citado possa ser localizado e recuperado na mesma condição, ou o mais próximo possível, em perpetuidade. Os dados publicados devem ser mantidos por uma instituição que forneça curadoria adequada em longo prazo. A publicação de dados deve ter referência inequívoca e persistente;
- c) garantia de que haverá *software* capaz de interpretar o objeto quando for recuperado. A curadoria deve tratar as questões de migração das interfaces e *softwares*.

Apesar de oneroso, o processo ativo de curadoria é necessário para a publicação de dados (LAWRENCE *et al.*, 2011; PARSONS; FOX, 2013).

A seguir, apresentamos os modelos de publicação de dados disponíveis.

2.7.1. Modelos de publicação de dados

As tentativas para publicação de dados podem ser agrupadas em três categorias (KRATZ; STRASSER, 2014):

- a) dados que suplementam um documento de pesquisa tradicional. Nesta categoria, um artigo de periódico tradicional é acompanhado de dados adicionais, que podem ser

¹² Uniform Resource Locator

¹³ Digital Object Identifier

hospedados pelo periódico como material suplementar ou depositado em repositório de terceiros como o Dryad¹⁴, Figshare¹⁵, Zenodo¹⁶, repositórios específicos por disciplina com o GenBank¹⁷ ou repositórios institucionais;

- b) dados independentes de vinculação a um artigo. O conjunto de dados é acompanhado de informações descritivas (metadados) e são depositados em repositórios sem assumir a forma de artigo de revista;
- c) dados em forma de um ‘artigo de dados’. Um *data paper* descreve um conjunto de dados de forma detalhada, bem como o método de coleta utilizado sem fazer qualquer análise ou conclusão.

As primeiras tentativas para suporte à publicação de dados (*data publication*) foram a inclusão dos dados como parte integrante do artigo e a disponibilização de dados em arquivos suplementares anexados ao artigo principal. Nesses modelos de publicação, os leitores não podem encontrar e vincular dados de forma independente da publicação principal (CANDELA *et al.*, 2015).

Os problemas apresentados expõem a necessidade de se estabelecer um novo modelo de publicação baseado no conceito de ‘artigo de dados’ (*data paper*). O artigo de dados descreve o conjunto de dados, seus métodos de raciocínio e coleta, contudo, sem oferecer análise ou conclusão (KRATZ; STRASSER, 2015).

Os *data papers* foram anunciados em editorial da revista *The International Journal of Robotics Research* em 2009 (CANDELA *et al.*, 2015; NEWMAN; CORKE, 2009). Representam novo gênero de artigo de periódico. Recebem tratamento similar ao de artigos tradicionais, passando pelo processo padrão de revisão por pares e impressos – ou disponibilizados *online* - em periódicos, passíveis de citação como os artigos tradicionais (NEWMAN; CORKE, 2009).

Os *data journals* foram anunciados em 2011 por Pfeiffenberger e Carlson (2011). São periódicos voltados especificamente para a publicação de *data papers*, que permitem a recompensa aos autores assim como as publicações tradicionais (CANDELA *et al.*, 2015; PFEIFFENBERGER; CARLSON, 2011).

O processo de revisão por pares de dados

¹⁴ <http://datadryad.org/>

¹⁵ <https://figshare.com/>

¹⁶ <https://zenodo.org/>

¹⁷ <https://www.ncbi.nlm.nih.gov/genbank/>

deve garantir que todos os metadados sejam tão completos quanto possível, mas também deve abordar outras qualidades esperadas do material [...] como a consistência interna dos dados, o mérito dos algoritmos usados, a importância dos dados e seu impacto potencial (LAWRENCE *et al.*, 2011, p. 11, tradução nossa).

Um conjunto de dados que passou por revisão por pares pode ser compreendido como processo de garantia, que monitora a qualidade científica (LAWRENCE *et al.*, 2011).

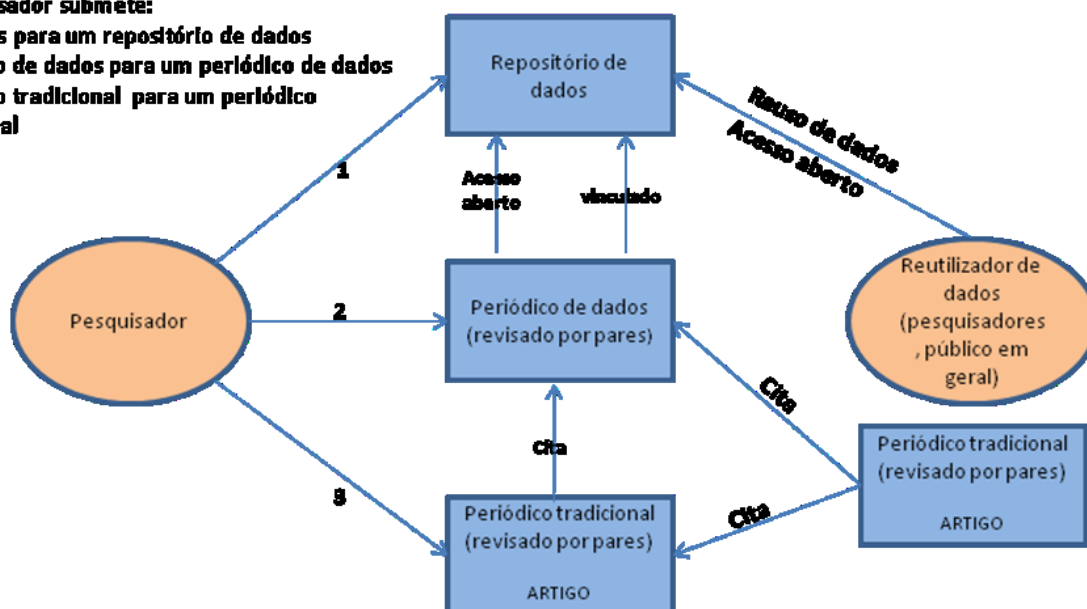
Pfeiffenberger e Carlson (2011) lembram que publicação do conjunto de dados na forma de artigo de dados em periódicos não substituirá os repositórios. Os conjuntos de dados devem ser depositados em repositórios confiáveis, devendo estabilidade de conexão, como, por exemplo, o uso do *Digital Object Identifier* (DOI).

Segundo Pfeiffenberger e Carlson (2011), o pesquisador disponibiliza seu conjunto de dados em um repositório de dados (*data repository*) e submete os artigos, em formato tradicional e em formato de artigo de dados (*data papers*) aos periódicos. Estes são revisados por pares – periódicos tradicionais e periódicos de dados (*data journals*), garantindo a recompensa por meio das citações, tanto dos artigos tradicionais quanto dos artigos de dados. Esses últimos são vinculados diretamente aos conjuntos de dados disponibilizados nos repositórios. O depósito em repositórios garante o reuso de dados, conforme apresentado na FIG. 8.

FIGURA 8 - Locais ideais de submissão e citação com o uso dos periódicos de dados

O pesquisador submete:

- 1 – Dados para um repositório de dados
- 2 – Artigo de dados para um periódico de dados
- 3 – Artigo tradicional para um periódico tradicional



Fonte: Adaptado de Pfeiffenberger e Carlson (2011, p. 8).

Nem sempre a publicação de dados garante a citação e o devido crédito acadêmico ao pesquisador (KRATZ; STRASSER, 2015). Mas, cumprindo as funções básicas propostas

por Van de Sompel *et al.* (2004) e Lawrence *et al.* (2011), a publicação de dados torna-se incentivo para documentar e compartilhar dados (KRATZ; STRASSER, 2015).

Dessa forma, a comunidade científica tem adotado a publicação de dados (com P maiúsculo) como incentivo à documentação e ao compartilhamento, pois a proposta da publicação de dados é a apropriação “[...] do prestígio da autoria na literatura revisada por pares para recompensar os pesquisadores que criam conjuntos de dados úteis e bem documentados” (KRATZ; STRASSER, 2015, p.1, tradução nossa). De acordo com os autores, a publicação de dados apropria-se de uma terminologia familiar, a publicação e a revisão por pares, colocando os dados no sistema de recompensa acadêmica já existente.

2.8. CITAÇÃO DE DADOS

Ainda na seara da publicação de dados, outro aspecto importante a considerar é a citação de dados. Para o *Data Citation Synthesis Group* (2014, *online*, tradução nossa), “a citação de dados, como a citação de outras evidências e fontes, é boa prática de pesquisa e faz parte do ecossistema acadêmico que apoia o reuso de dados”. O *Data Thesaurus* complementa que “uma citação de dados fornece a atribuição de fontes de dados facilitando o acesso a dados de pesquisa em periódicos e na *Internet*” (NNLM, 2018, *online*, tradução nossa).

A citação de dados traz como aspecto importante o “[...] desejo de fornecer crédito justo para o esforço intelectual e técnico que envolve a criação de um bom conjunto de dados [...],” sendo vista como incentivo para a liberação de dados (PARSONS; FOX, 2013, p. 39, tradução nossa). Além disso, a citação de dados pode auxiliar o reuso e a verificação de dados ao possibilitar o rastreamento dos responsáveis pela sua criação para que eles sejam verificados e reutilizados em estudos futuros (NNLM, 2018).

Para que haja citação, [...] os dados precisam ser identificados e localizados, de preferência, com identificador persistente, tão logo estiverem disponíveis para uso por qualquer pessoa que não seja o criador original (PARSONS; FOX, 2013, p. 39, tradução nossa).

Os identificadores persistentes “[...] podem ser citados para recuperar objetos digitais durante um longo período de tempo” (IFLA, 2017). Fornecem meio de identificação única e persistente dos objetos digitais. A atribuição de um identificador persistente aos conjuntos de dados ajuda na localização desses conjuntos face à vasta quantidade de dados gerados. Os identificadores persistentes são amplamente utilizados no ambiente digital e podem fornecer informações sobre o objeto, independentemente de onde esteja localizado (NNLM,

2018). Os identificadores persistentes incluem DOIs¹⁸, ARKs¹⁹, Handles²⁰ e ORCID²¹ (NNLM, 2018).

A *Earth Science Information Partners* (ESIP) (2012) acrescenta que a citação de dados é prática científica em evolução que tem como propósito primordial auxiliar a reprodutibilidade científica através de referência direta e inequívoca aos dados precisos utilizados em um estudo particular. A citação de dados apresenta ainda outros propósitos como (ESIP, 2012):

- a) fornecer crédito justo para criadores de dados ou autores;
- b) garantir transparência científica e responsabilidade razoável para autores e administradores;
- c) auxiliar no acompanhamento do impacto do conjunto de dados por meio de referências na literatura científica;
- d) ajudar os produtores a verificarem como seus dados estão sendo usados;
- e) ajudar futuros usuários a identificarem como outros pesquisadores usaram os dados.

No entanto, a ESIP (2012) adverte que, apesar dos benefícios, os dados são raramente citados. Há diversas razões para que isso ocorra. Uma delas é a “[...] falta de recomendações consistentes sobre como citar e referenciar um conjunto de dados e como construir efetivamente uma citação de dados adequada” (ESIP, 2012, p. 1, tradução nossa).

¹⁸ O DOI (*Digital Object Identifier*) é um código alfanumérico individual usado para identificação persistente e recuperação mais rápida e precisa de publicações científicas como artigos, trabalhos de eventos e outros documentos em formato digital (IDF, 2018), inclusive conjunto de dados. O DataCite é uma organização sem fins lucrativos que fornece identificadores persistentes (DOIs) para dados de pesquisa (DATACITE, 2019).

¹⁹ O identificador ARK é um esquema de nomenclatura para acesso persistente a objetos digitais (incluindo imagens, textos, conjuntos de dados e recursos de pesquisa), desenvolvido em 2001 e implementado pela Biblioteca Digital da Califórnia (IFLA, 2017). O identificador ARK (*Archival Resource Key*) é um *Uniform Resource Locator* (URL) projetado para suportar acesso de longo prazo a objetos de informação de qualquer tipo como objetos digitais - documentos, bases de dados, imagens, *software*, *sites etc.*; objetos físicos - livros, ossos, estátuas *etc.*; seres vivos e grupos - pessoas, animais, empresas, orquestras *etc.*; objetos intangíveis - lugares, produtos químicos, doenças, termos de vocabulário, performances *etc.* (UC, 2019).

²⁰ O *Handle System*, desenvolvido pela *Corporation for National Research Initiatives* (CNRI), foi projetado como sistema de resolução para objetos digitais. O *Handle System* permite, ainda, atribuir, gerenciar e resolver identificadores persistentes, conhecidos como ‘*handles*’ (IDF, 2018).

²¹ O ORCID ID é um identificador digital persistente para pesquisadores, gerenciado e fornecido pela organização sem fins lucrativos ORCID (*Open Researcher and Contributor ID*) (NNLM, 2018; ORCID, 2019). O ORCID ID “[...] permite tanto a desambiguação do pesquisador quanto um método de associar precisamente os autores a todas as suas obras, independentemente de como seu nome e afiliações institucionais mudam com o tempo” (NNLM, 2018, *online*, tradução nossa).

Tal fato confirma o que fora identificado por Lawrence *et al.* (2011), que afirmam que não há uma convergência de sintaxe, uma maneira padrão de se referir a um conjunto de dados.

Desse modo, apesar da falta de recomendações consistentes sobre como citar e referenciar um conjunto de dados (ESIP, 2012), o documento *Joint Declaration of Data Citation Principles* do *Data Citation Synthesis Group* (2014) incentiva as comunidades a desenvolverem práticas e ferramentas de citação que incorporem os seguintes princípios:

- a) importância - as citações de dados devem ter a mesma importância no registro acadêmico que as citações de outros objetos de pesquisa;
- b) crédito e atribuição - as citações devem facilitar a concessão de crédito acadêmico e atribuição normativa e legal a todos os contribuintes dos dados;
- c) evidência - na literatura científica, sempre que uma afirmação se basear em dados, os dados correspondentes devem ser citados;
- d) identificação única – a citação de dados deve incluir um método persistente de identificação, acionável por máquina, globalmente exclusivo e amplamente utilizado por uma comunidade;
- e) acesso - as citações de dados devem facilitar o acesso aos dados em si e a esses metadados, documentação, código e outros materiais associados;
- f) persistência - identificadores únicos e metadados descrevendo os dados devem ser persistentes, mesmo além do tempo de vida dos dados que descrevem;
- g) especificidade e verificabilidade – as citações devem facilitar a identificação, o acesso e a verificação dos dados específicos que apoiam uma reivindicação;
- h) interoperabilidade e flexibilidade – os métodos de citação devem ser flexíveis sem, contudo diferir de forma que comprometa a interoperabilidade das práticas de citação entre as comunidades.

Em síntese, a publicação de dados busca definir conjuntos de dados bem descritos, com certo nível de qualidade ou revisão por pares. Assim, “os dados publicados são então considerados artefatos científicos de primeira classe, capazes de referência [...]” (PARSONS; FOX, 2013, p. 34, tradução nossa) ou seja, citados. A citação de dados adquire, dessa forma, ênfase importante na publicação de dados (PARSONS; FOX, 2013).

2.9. INFRAESTRUTURAS, PADRÕES E POLÍTICAS PARA GESTÃO DE DADOS

Como já abordado anteriormente, a gestão de recursos digitais envolve preservação e curadoria. Além dos pesquisadores capazes de documentar com precisão os seus dados, uma boa gestão de dados requer, das instituições, a garantia que os processos de gestão sejam realizados da melhor forma possível (WILSON *et al.*, 2010).

Sayão e Sales (2016) afirmam que, para que as coleções de dados de pesquisa sejam reutilizadas, é necessário um conjunto de ações específicas que permeiem todo o ciclo de vida dos dados. Essas ações iniciam

no planejamento da criação dos dados, passam pela organização desses dados em coleções identificadas por meio de referências estáveis e padronizadas e vão até o arquivamento de longo prazo para os dados reconhecidos como de valor permanente ou que são referências universais (SAYÃO; SALES, 2016, p. 91).

Essas ações são realizadas por meio de uma infraestrutura de dados - uma “[...] orquestração de tecnologias, padrões de dados e metadados e políticas embutidas no empreendimento de pesquisa”, podendo existir em escalas nacionais ou internacionais, dentro de uma instituição ou comunidade de pesquisa (QIN, 2013, p. 216, tradução nossa).

As infraestruturas asseguram que a gestão de dados seja realizada de maneira adequada dentro das instituições, garantindo a preservação ativa e em longo prazo dos dados de pesquisa gerados e coletados, “[...] mantendo as suas características de autenticidade integridade e proveniência, de forma que eles estejam sempre disponíveis e prontos para serem usados” (SAYÃO; SALES, 2015, p. 23-24). Contudo, a gestão dos dados, bem como a infraestrutura necessária, variam de acordo com a escala do projeto científico e seu grau de homogeneidade e heterogeneidade (BORGMAN *et al.*, 2015).

Na *big science*, as disciplinas desenvolveram suas próprias ferramentas para armazenar, pesquisar e acessar dados relevantes para a área (ONSRUD; CAMPBELL, 2007). Na *long tail science* o cenário está mudando pelo desenvolvimento de infraestruturas e na tentativa de conciliar padrões, formatos, metadados e documentação, garantindo a compatibilidade de sistemas, maximização das descobertas e ligação entre os conjuntos de dados (KITCHIN; LAURIAULT, 2015).

Assim, “à medida que o volume de dados da cauda longa aumenta, a comunidade científica está desenvolvendo e mantendo Ciberinfraestruturas para gerenciar e compartilhar seus dados”. Os autores citam como exemplo o *Consortium of Universities for the Advancement of Hydrologic Science* (CUAHSI) (<https://www.cuahsi.org/>); *DataOne* (<https://www.dataone.org/>); *Sustainable Environmental Actionable Data* (SEAD) (<http://sead-data.net/>); e *Integrated Data Management System for Critical Zone Observatories* (CZOData) (<http://criticalzone.org/national/data/cyberinfrastructure-development-1national/>). (ELAG *et al.*, 2017, p. 100, tradução nossa).

Admitindo a mesma ideia de Qin (2013), Rodrigues *et al.* (2010) afirmam que, além das infraestruturas (serviços, sistemas, normas e protocolos) necessárias para garantir a coleta, preservação e acesso aos dados, há outros requisitos na gestão de dados científicos, quais

sejam: os aspectos políticos, legais e éticos associados ao acesso e reutilização dos dados científicos além do contexto inicial em que foram criados ou coletados.

No mesmo sentido, Friedlander e Adler (2006) afirmam que as infraestruturas vão além das instalações físicas e ferramentas de *software*, devendo:

- suportar múltiplas representações de dados por diversos usuários, agora e no futuro, representando diversas culturas disciplinares;
- suportar novas camadas de serviço e reutilizar componentes através de modelos de dados e metadados, bem como políticas institucionais que regem o uso de coleções e serviços;
- suportar sistemas de segurança apropriados, privacidade e confidencialidade; e
- ativar relacionamentos confiáveis em vários níveis (FRIEDLANDER; ADLER, 2006, p. 28, tradução nossa).

Já Sayão e Sales (2013) indicam outras instâncias a serem consideradas: infraestrutura organizacional, desenvolvimento de coleções de dados, pesquisa, formação de recursos humanos, sustentabilidade econômica, implicações sociais, legais e éticas e disponibilização de serviços.

As infraestruturas devem ser amplas e incluir: a política da universidade; sistemas para preservação e documentação de dados de pesquisa; treinamento e suporte; ferramentas de *software* para a visualização de imagens grandes; e criação e compartilhamento de bases de dados via *web* (WILSON *et al.*, 2010; WILSON *et al.*, 2011).

Qin (2013) argumenta que não existe uma definição única e conceitual para infraestrutura de dados. Entretanto, a autora traduz de maneira didática o conceito de infraestrutura de dados em três dimensões: 1) tecnologias; 2) políticas que governam o gerenciamento, compartilhamento e uso de dados; 3) padrões de dados e metadados.

1. A dimensão tecnológica abrange uma série de tecnologias para coletar, armazenar, processar, organizar, transmitir e preservar dados e, ainda, plataformas para comunicação e colaboração. Nessa dimensão tecnológica estão incluídas as redes, bancos de dados, sistemas de autenticação e aplicativos de *software*.
2. As políticas que governam o gerenciamento, compartilhamento e uso de dados – nacionais ou globais. Estão, principalmente, relacionadas ao compartilhamento de dados, proteção à propriedade intelectual, questões éticas e acesso aberto. As políticas nacionais e institucionais fornecem a estrutura para que pesquisadores e projetos individuais estabeleçam suas próprias políticas no dia-a-dia da operação. A natureza dos dados pode afetar diretamente a política. Dados obtidos a partir de observação de fenômenos naturais que não podem ser reproduzidos ou replicados

deverão ser preservados. Os dados gerados a partir de ensaios clínicos devem ser replicados. Portanto, não precisam ser preservados.

3. Os padrões de dados e metadados representam a complexidade dos tipos e formatos de dados resultando em grande número de padrões tanto para dados quanto para metadados. Contudo, são necessários formatos de dados padronizados, permitindo que os dados sejam: a) conversíveis; b) lidos em qualquer lugar; c) expansíveis (adicionar tipos e estruturas posteriormente). Além disso, são necessários padrões de metadados, definindo os elementos e as estruturas usadas para descrever os conjuntos de dados (coleta - quem, onde, como e quais tipos), sendo essenciais para a descoberta e uso efetivo dos dados.

Há uma diversidade de requisitos para infraestrutura de gestão de dados em diferentes áreas de pesquisa. Esses requisitos devem ser específicos para cada área ou projeto. Os autores explicam que alguns podem se concentrar em grandes recursos de armazenamento (*terabytes* de dados), outros podem se preocupar em manter seus dados confidenciais seguros. Há outros que focam no compartilhando de dados com outros pesquisadores ou público em geral. Apesar de específicos, há aspectos comuns exigidos em todas as áreas ou projetos como: a documentação de dados e metadados; treinamento e suporte; armazenamento seguro; e vinculação de dados a publicações (WILSON *et al.*, 2010).

Onsrud e Campbell (2007) afirmam que embora a infraestrutura possa variar, elas terão que lidar com pelo menos cinco questões críticas:

- a) gestão da propriedade intelectual – licenciamento;
- b) geração de metadados;
- c) rastreamento de origem;
- d) arquivamento;
- e) avaliação por pares.

Para além de todas essas questões, as infraestruturas devem ser simples, tanto para o pesquisador que disponibiliza os dados quanto para os usuários e possíveis reutilizadores (ONSRUD; CAMPBELL, 2007).

2.9.1. Infraestrutura tecnológica

É muito comum falar em ciberinfraestrutura - uma infraestrutura abrangente que atende a atividades de pesquisa amplamente distribuídas e, muitas vezes, em grande escala, conectando de forma eficiente laboratórios, dados, computadores e pessoas (SALES; SAYÃO, 2015) - nas áreas da *big science*. Entretanto, as necessidades de infraestrutura de dados na *long tail science* são diferentes (GOLD, 2007b), uma vez que, segundo apontam Rodrigues *et al.* (2010)

em algumas áreas [principalmente na áreas da *big science*] existem centros de dados que podem assegurar uma curadoria profissional de dados considerados relevantes enquanto o volume de dados que processam não ultrapassar a sua capacidade. O problema é o que fazer com a “pequena ciência” [ou *long tail science*], isto é, com os conjuntos de dados produzidos por investigadores individualmente ou por pequenos grupos de investigação que não possuem nem os recursos nem as infra-estruturas para cuidar dos conjuntos de dados para além do termo dos seus projectos de investigação (RODRIGUES *et al.*, 2010, p. 17).

Um componente essencial dessas infraestruturas são as bibliotecas digitais. Elas podem ser implantadas “[...] ao longo do ciclo de vida da pesquisa, desde capturar observações através de limpeza, análise e interpretação de dados, gerenciamento, conservação e administração de produtos de pesquisa” (BORGMAN *et al.*, 2015, p. 208, tradução nossa), sendo vistas como alternativa de sistemas locais para gerenciamento dos dados de pesquisa.

As bibliotecas digitais, tanto para dados ou documentos, são normalmente implementadas no final do processo de pesquisa visando fornecer acesso a publicações e dados para fins de reutilização, verificação ou reprodutibilidade (BORGMAN *et al.*, 2015).

A ideia do uso de bibliotecas digitais em uma concepção mais ampla, que suporte todo o ciclo de vida da informação, não é nova. A noção mais inclusiva, abarcando todo o ciclo de vida, requer uma arquitetura diferente daquelas bibliotecas digitais voltadas somente para documentos – repositórios institucionais tradicionais -, uma vez que os dados “[...] são mais bem vistos do que papéis, protocolos, ferramentas analíticas, instrumentos, *software*, fluxos de trabalho e outros componentes da prática de pesquisa” (BORGMAN *et al.*, 2015, p. 209, tradução nossa).

Em pesquisas realizadas dentro das universidades, portanto, no âmbito da *long tail science*, os repositórios são considerados parte central da infraestrutura, ou uma infraestrutura adequada para a gestão de dados científicos (BRANDT, 2007; CORRALL, 2012; COX; PINFIELD, 2014; CRAGIN *et al.*, 2010; GOLD, 2007a; HEY; HEY, 2006; LEE; STVILIA, 2017; LUCE, 2008; NEWTON; MILLER; BRACKE, 2010; POOLE, 2016; RODRIGUES *et al.*, 2010; SAYÃO; SALES, 2016; TENOPIR *et al.*, 2013, 2015, 2017; TENOPIR; BIRCH; ALLARD, 2012). Apoiam a gestão de dados de pesquisa em todo o seu ciclo (SAYÃO; SALES, 2016).

2.9.1.1. Repositórios digitais

Para o Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT) (2012, *online*), os repositórios digitais “[...] são bases de dados *online* que reúnem de maneira organizada a produção científica de uma instituição ou área temática”.

De acordo com Rodríguez Bravo e Alvite Díez (2007, p. 365, tradução nossa), os repositórios digitais cresceram sob o auspício das bibliotecas e têm como objetivos “[...] incentivar a difusão do patrimônio artístico, bibliográfico e documental, do qual essas instituições são as guardiãs”.

A literatura apresenta diversos tipos de repositórios digitais (SALES, 2014), conforme apresentado no Quadro 12.

QUADRO 12 - Classificação de repositórios

Classificação	Autor	Exemplos
Institucional	(LEE; STVILIA, 2017; LEITE, 2009; SALES, 2014)	Repositório Institucional da UFMG (https://repositorio.ufmg.br/) Lume – Repositório Digital da UFRGS (https://lume.ufrgs.br/)
Disciplinar ou temático	(LEE; STVILIA, 2017; LEITE, 2009; SALES, 2014)	Cristalografia - Crystallography Open Database (http://www.crystallography.net/cod/)
Domínio	(LEE; STVILIA, 2017)	Química - NIST Chemistry WebBook (https://webbook.nist.gov/chemistry/)
Teses e dissertações	(LEITE, 2009)	Biblioteca Digital de Teses e Dissertações da USP (https://www.teses.usp.br/) Biblioteca Digital da PUC Minas (https://web.sistemas.pucminas.br/BDP/PUC%20Minas)

Fonte: Elaborado pela autora.

Lee e Stvila (2017) consideram três tipos de repositórios: de domínio; disciplinar e; institucional. Já Leite (2009) aponta que os repositórios digitais podem ser de três tipos: repositórios institucionais; repositórios temáticos ou disciplinares e; repositórios de teses e dissertações. Sales (2014) aponta que uma classificação mais comum divide os repositórios digitais apenas em institucionais e temáticos. Esses dois tipos de repositórios serão tratados a seguir.

Os repositórios institucionais lidam com a produção científica de uma determinada instituição (IBICT, 2012) e tem como objetivo colecionar, preservar e dar visibilidade à produção científica, técnica ou administrativa de uma instituição. Além disso, cumprem a função de auditoria das instituições (RODRIGUES *et al.*, 2010).

De acordo com Leite (2009, p. 20), os repositórios institucionais são “[...] voltados à produção intelectual de uma instituição, especialmente universidades e institutos de pesquisa”. Assim, os repositórios institucionais mantidos pelas universidades podem ser vistos como

um conjunto de serviços que uma universidade oferece aos membros de sua comunidade para a gestão e disseminação de materiais digitais criados pela instituição e seus membros da comunidade (LYNCH, 2003, *online*, tradução nossa).

O repositório institucional é apoiado por um conjunto de tecnologias da informação (LYNCH, 2003). De acordo com Powell e Heery (2006, p. 7, tradução nossa), as tecnologias e plataformas de *software* evoluirão. Por isso, é importante que a ênfase seja dada aos “[...] “serviços de repositório”, e não no repositório como plataforma de *software* específica”.

A manutenção de um repositório institucional representa compromisso organizacional com o gerenciamento dos materiais digitais produzidos pela instituição, com a preservação em longo prazo e com a organização e acesso e distribuição. O repositório institucional não é simplesmente um conjunto fixo de *software* e *hardware*, pois a gestão de mudanças tecnológicas e a migração de conteúdo digital é parte fundamental dos serviços de repositórios (LYNCH, 2003).

Os repositórios disciplinares ou temáticos lidam com a produção científica de uma determinada área, sem impor limites institucionais (IBICT, 2012). Segundo Leite (2009) esses repositórios são voltados às comunidades científicas específicas, tratando, portanto, da produção intelectual de uma área do conhecimento em particular.

Em se tratando das universidades, as

[...] comunidades de pesquisa tendem a ser heterogêneas nos métodos e tipos de dados aplicados, sem padrões de dados uniformes ou amplamente aplicados, e atualmente não são bem suportados pelos serviços de repositório disciplinar (CRAGIN *et al.*, 2010, p. 4025, tradução nossa).

Os repositórios institucionais seriam mais indicados e abarcaria também as coleções de dados além dos documentos tradicionais. Precisarão acomodar ampla gama de características de dados subdisciplinares e práticas de dados, incluindo as práticas de compartilhamento (CRAGIN *et al.*, 2010), desempenhando papel cada vez mais importante no ciclo de vida dos dados (HENTY, 2008).

Independentemente do tipo ou classificação do repositório, a responsabilidade operacional pelos serviços de repositórios é dividida entre: bibliotecários; tecnólogos de informação; administradores de arquivos e registros; professores e administradores de universidades; e formuladores de políticas (LYNCH, 2003).

Além das classificações de repositórios apresentadas no Quadro 12, Sales (2014) admite trabalhar com uma classificação segundo o tipo de material depositado. Dessa forma, os repositórios podem ser: repositórios de *e-prints* (versão eletrônica de teses, artigos, relatórios); repositórios de dados de pesquisa; ou outro tipo de material.

2.9.1.2. Repositórios de dados de pesquisa

Os pesquisadores precisam, por um lado, de “[...] infraestruturas que assegurem o máximo de confiabilidade, estabilidade e acessibilidade e que facilitem o trabalho de arquivamento, compartilhamento e reconhecimento de autoria para os seus dados [...]” (SAYÃO; SALES, 2016, p. 93). Por outro lado, também necessitam

[...] encontrar coleções de dados de pesquisa, saber como acessá-las e sob que condições podem reutilizar esses dados e assim dar prosseguimento às suas pesquisas confiando na autenticidade e proveniência dos dados coletados ou gerados por outros pesquisadores (SAYÃO; SALES, 2016, p. 93).

As instituições começaram a expandir seus serviços de repositórios digitais, diante das exigências de pesquisadores, agências de fomento e da sociedade. Os repositórios digitais, institucionais ou temáticos, há pouco tempo voltados somente para publicações, agora recebem o depósito de dados de pesquisa (SAYÃO; SALES, 2016).

Lynch (2003) e Powell e Heery (2006) já apontavam que os repositórios institucionais também abrigariam os dados de pesquisa²². Todavia, Sayão e Sales (2017, p. 68) acreditam que a melhor opção é o depósito dos dados em repositórios voltados, especificamente, para dados, “[...] sejam institucionais, disciplinares, multidisciplinares, orientados para projetos ou os ligados aos periódicos científicos”.

Os repositórios de dados²³ (*data repositories*), também conhecidos como repositórios de dados de pesquisa (*research data repositories*) ou repositórios de dados científicos (*scientific data repositories*) ou ainda *data centers*, são considerados parte importante da infraestrutura tecnológica para o compartilhamento e reutilização de dados (CURTY, 2015).

De acordo com Curty (2015, p. 13, tradução nossa), os repositórios de dados podem ser vistos como “[...] plataformas institucionais, interinstitucionais, disciplinares ou multidisciplinares baseadas na *web* responsáveis pelo armazenamento, organização, curadoria e preservação de dados de pesquisa para acesso de longo prazo”, oferecendo um conjunto mínimo de serviços básicos como autoarquivamento, pesquisa, acesso e controle de proveniência.

²² Exemplos de repositórios institucionais que abrigam dados de pesquisa: *University of Southampton Institutional Research Repository* (<https://eprints.soton.ac.uk/>); IDEALS (<https://www.ideals.illinois.edu/>), dentre outros.

²³ Exemplos de repositórios de dados de pesquisa: *PPBio Data Repository* (<https://ppbiodata.inpa.gov.br/metacatui/>); *European MassBank* (<https://massbank.eu/MassBank/>); *INDEPTH Data Repository* (<http://www.indepth-ishare.org/index.php/home>); *LSHTM Data Compass* (<https://datacompass.lshtm.ac.uk/>); *Research Data Leeds Repository* (<http://archive.researchdata.leeds.ac.uk/>). Outros exemplos podem ser encontrados no *Registry of Research Data Repositories* (<https://www.re3data.org/>).

Dessa forma, os repositórios de dados são considerados por Sayão e Sales (2015, p. 50) como o local “[...] mais apropriado para que seus [os] dados sejam preservados e possam ser recuperados, acessados e citados por outros pesquisadores, ou seja, tenham visibilidade em escala mundial”. Logo, o uso de repositórios de dados traz como benefícios (SAYÃO; SALES, 2017b):

- a) maior visibilidade dos dados, garantindo mais citações;
- b) maior grau de reuso e compartilhamento;
- c) estruturas de preservação e armazenamento seguro;
- d) possibilidade de interação entre os pesquisadores, anotações e administração de licenças e de tempo de embargo;
- e) estabelecimento de *links* entre dados e artigos de periódicos.

Powell e Heery (2006) apresentam algumas questões que devem ser desenvolvidas na área de repositórios como:

- a) investimento em repositórios de dados de pesquisa;
- b) políticas robustas de aquisição e mecanismos de descoberta de dados;
- c) acordos sobre identificadores e citações exclusivas para todos os conjuntos de dados;
- d) funcionalidades e serviços que suportem curadoria, migração e preservação;
- e) interação *online* com dados de pesquisa e *links* para todas as informações relacionadas de todos os tipos de mídias;
- f) gerenciamento adequado dos direitos de propriedade intelectual em dados de pesquisa;
- g) trabalho em conjunto para dar suporte aos repositórios de dados de pesquisa (comunidades de TI, dados e informações);
- h) desenvolvimento de mecanismos que fomentem advocacia, mandatos, financiamento, adotantes antecipados e demonstradores.

Powell e Heery (2006, p. 16, tradução nossa) afirmam que é necessária uma infraestrutura técnica comum para repositórios, garantindo a interoperabilidade entre todos os seus tipos, aumentando a visibilidade da base de evidências (publicações e dados), além de encorajar a citação e reutilização de dados de pesquisa.

Embora Powell e Heery (2006) tenham projetado o cenário para o ano de 2010, muitas questões apontadas ainda requerem empenho por parte das instituições, representando desafios para a gestão de dados de pesquisa realizada pelas instituições. Os repositórios terão que se adaptar a cada disciplina, pois “[...] uma abordagem genérica à curadoria de

dados não será suficiente para lidar com todas as necessidades e expectativas dos investigadores das diferentes áreas (RODRIGUES *et al.*, 2010, p. 18).

2.9.2. Infraestrutura política

Um grande número de políticas relacionadas a dados vem sendo mundialmente adotado por agências de financiamento e instituições com objetivo de melhorar a eficiência da pesquisa, a colaboração e a transparência (LYON, 2007; SHEARER, 2015).

Dessa forma, tanto agências de financiamento quanto as instituições são responsáveis pelas políticas. Lyon (2007, p. 6, tradução nossa) afirma que cada agência de financiamento “[...] deve publicar, implementar e impor abertamente uma Política de Gerenciamento, Preservação e Compartilhamento de Dados”. Para a autora, as agências financiadoras

devem tornar suas políticas claras e consistentes e torná-las acessíveis e apropriadas ao público pretendido. Essas políticas devem abranger gestão e curadoria de dados, publicação e compartilhamento e preservação de dados. Devem levar em conta a paisagem em constante mudança de *data centers* e arquivos de dados, repositórios institucionais e de assunto, *software* de rede social, como *wikis* e *blogs*, e outras oportunidades de publicação na *web* (LYON, 2007, p. 46, tradução nossa).

Por outro lado, as instituições precisam de política de gerenciamento, curadoria, compartilhamento e preservação de dados. Assim,

cada instituição de ensino superior deve implementar uma Política de Gestão, Preservação e Compartilhamento de Dados, que recomende o depósito de dados em repositório de dados de acesso aberto apropriado e/ou *data center* onde eles existam (LYON, 2007, p. 6, tradução nossa).

As políticas, de forma geral, preveem como os dados serão gerenciados em todo o ciclo de vida, de forma a possibilitar “[...] a gestão de dados e permitir que sejam mais amplamente usados” (SHEARER, 2015, p. 7, tradução nossa).

De acordo com Shearer (2015, p. 4, tradução nossa), um elemento importante no contexto das políticas são os planos de gestão de dados, “[...] pois obrigam os pesquisadores a pensarem sobre como eles administrarão seus dados antes do projeto, um requisito essencial para boas práticas de gestão de dados”.

As políticas de gestão de dados de pesquisa estão situadas no contexto de um conjunto mais amplo de princípios e objetivos que orientam seus requisitos específicos. Em geral, as políticas permitem: acelerar a pesquisa; apoiar novos *insights* e descobertas; promover a colaboração; melhorar a eficiência da pesquisa e; facilitar a prestação de contas (SHEARER, 2015).

Os componentes mais frequentes das políticas são os padrões de metadados,

compartilhamento e retenção de dados e/ou preservação de longo prazo. As políticas contêm disposições para a proteção da confidencialidade, propriedade intelectual e dados confidenciais (SHEARER, 2015).

Dessa forma, Shearer (2015) enumera os elementos mais comuns encontrados numa política de gestão de dados de pesquisa (Quadro 13):

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

Requisitos da política	
Qualidade e padrões de dados	Os investigadores são obrigados a aderir aos padrões internacionais para permitir o acesso e a reutilização.
	Documentação de dados e metadados devem 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 determinado período de tempo mínimo.
	Quando possível, os investigadores devem depositar seus dados em arquivo de longo prazo para garantir a preservação de seus dados.
Planos de gestão de dados	As propostas de pesquisa devem incluir Plano de Gestão de Dados.
Disposições comuns às políticas	
Privacidade	Os direitos e 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.
Outros aspectos	
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, tradução nossa).

Existe uma diversidade em termos de como as políticas são monitoradas ou implementadas. Embora as políticas possam ser adotadas isoladamente, Shearer (2015) admite que só elas não garantem uma boa gestão de dados. As boas práticas são amparadas em diversos fatores como incentivos, especialização, serviços e infraestrutura, bem como mecanismos de financiamento apropriados.

2.9.3. Padrões de dados e metadados

Os padrões de dados são um conjunto acordado de regras (NLM, 2013), políticas ou práticas recomendadas que definem como os dados devem ser formatados, bem como os metadados e documentação necessárias (NNLM, 2018), permitindo “[...] que as informações sejam compartilhadas e processadas de maneira uniforme e consistente” (NLM, 2013, *online*, tradução nossa).

Os padrões de dados incluem a representação, formato, definição, estruturação, marcação, transmissão, manipulação, uso e gestão de dados (USEPA, 2018), possibilitando a interoperabilidade²⁴ (ZHU; WU, 2011), a qualidade dos dados (USEPA, 2018; ZHU; LEE; ROSENTHAL, 2016) e a reutilização (NNLM, 2018).

Dessa maneira, o uso de padrões facilita “[...] a criação, o compartilhamento e a integração, garantindo que haja compreensão clara de como os dados são representados e quais recebidos estão no formato esperado” (USGS, 2018a, *online*, tradução nossa).

Criado a partir do consenso de especialistas, os padrões de dados trazem benefícios como: economia em longo prazo, pois o mesmo código e os mesmos mecanismos podem ser usados para fornecer: acesso à informação; transparência e compreensão. Promovem significados comuns e claros para dados que são reutilizados com frequência; facilidade de acesso aos dados; incentivo a reutilização de dados e *software*; resultados consistentes durante a recuperação de dados; comparação de dados, mesmo quando não são padronizados (USEPA, 2018). Os padrões tornam os dados mais úteis para além do projeto ou pessoa que os criou; e permitem a economia de tempo para conciliar quaisquer diferenças se as várias fontes concordarem com um padrão (USGS, 2018a).

Já os metadados “[...] são informações descritivas ou contextuais que se referem ou estão associados a outro objeto ou recurso”, sendo vistos como sustentação da curadoria digital,

²⁴ Interoperabilidade de dados “[...] é a capacidade de os dados serem analisados e/ou mesclados com dados semelhantes. A interoperabilidade de dados depende de padrões, documentação de dados e metadados para indicar aos pesquisadores quais conjuntos de dados ou variáveis são comparáveis” (NNLM, 2018, *online*, tradução nossa).

pois, sem eles, os recursos digitais poderiam se tornar irrecuperáveis, não identificáveis ou inutilizáveis (HIGGINS, 2007a, *online*).

Os metadados “[...] descrevem o recurso de informação e auxiliam na sua identificação, localização e recuperação, ao mesmo tempo em que facilita o gerenciamento de conteúdo e acesso” (HIGGINS, 2007b, *online*, tradução nossa). São compostos por vários elementos categorizados de acordo com as funções que eles suportam. Eles são essenciais não somente por permitir o acesso, mas por permitir a gestão efetiva dos dados (FRIEDLANDER; ADLER, 2006).

Segundo Starr (2003), o desejo de compartilhar informações foi motivador para o desenvolvimento e implementação de padrões de metadados, uma vez que as informações devem ser apresentadas em formato comum com o objetivo de serem compartilhadas.

Os padrões de metadados iniciam-se como esquemas desenvolvidos por uma comunidade específica de usuários, de acordo com suas necessidades, suportando algumas ou todas as funções (HIGGINS, 2007b) listadas a seguir:

- a) metadados descritivos – permitem a identificação, localização e recuperação de recursos de informação pelos usuários;
- b) metadados técnicos - descrevem os processos técnicos usados para produzir ou usar um objeto digital;
- c) metadados administrativos - gerenciam aspectos administrativos do objeto digital (direitos de propriedade intelectual e aquisição). Documentam informações referentes à criação, alteração e controle de versão dos próprios metadados;
- d) metadados de uso - gerenciam o acesso do usuário, o rastreamento do usuário e as informações sobre várias versões;
- e) metadados de preservação - documentam ações que foram realizadas para preservar um recurso digital, como migrações e cálculos de *checksum*²⁵.

O uso de padrões de metadados “[...] garante metadados ricos e consistentes que suportarão a descoberta, o uso e a integridade de recursos digitais em longo prazo” permitindo a interoperabilidade e maior potencial de descoberta de recursos (HIGGINS, 2007a, *online*, tradução nossa).

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

²⁵ Soma de verificação usada para verificar a integridade dos dados.

- a) 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;
- b) 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;
- c) padrões de marcação de metadados - garantem que os metadados sejam legíveis por máquina e que pesquisas automatizadas possam ser realizadas;
- d) padrões de empacotamento de metadados - definem os *links* entre objetos digitais e seus metadados enquanto vinculam os componentes em pacotes de arquivamento.

Existe um grande número de padrões de metadados (Apêndice A). Escolher qual padrão de metadados utilizar é tarefa complexa. Surgem questões como: qual padrão utilizar; a completude dos metadados; a consideração de custo *versus* benefício (quanto a instituição está disposta a investir) *etc.* Dessa forma, a avaliação e a preparação dos dados exigem um investimento significativo (NEWTON; MILLER; BRACKE, 2010).

Diante disso, ao escolher qual padrão implementar, deve-se considerar as seguintes questões: 1) contexto de criação e uso; 2) formato do material; 3) orçamento; 4) método de captura de metadados; e 5) armazenamento e entrega de metadados (HIGGINS, 2007a).

O contexto de criação e uso de metadados pressupõe o uso de um padrão que foi desenvolvido e testado em uma determinada comunidade. Deve-se, ainda, levar em consideração para quem, para qual comunidade os dados serão compartilhados e reusados.

O formato do material que está sendo descrito prevê o uso de padrões aplicáveis para formatos específicos como imagens, arquivos de áudio *etc.*

O orçamento presume criar os melhores dados possíveis que o recurso financeiro permitir, uma vez que fornecer pontos de acesso pode consumir grande parte do orçamento. Contudo, a funcionalidade e a capacidade de pesquisa serão significativamente aprimoradas.

O método de captura de metadados, que possibilita que os metadados possam ser gerados automaticamente ou a partir de entrada manual. Os metadados técnicos podem ser criados manualmente, enquanto os metadados descritivos geralmente requerem entrada manual.

O armazenamento e entrega tanto tem relação tanto com metadados quanto do recurso digital que eles descrevem. As opções de ferramentas de *software* terão impacto nas escolhas sobre os padrões de metadados selecionados.

Não é necessário usar todos os elementos do padrão escolhido. Há um conjunto de elementos obrigatórios e outros opcionais. Podem-se criar perfis de aplicação e extensões (Apêndice B) para uma determinada comunidade, assim como escolher o tipo de repositório ou tipos específicos de dados, de forma a garantir a informação mínima necessária à descrição adequada dos recursos, bem como a criação de índices e buscas e a implementação técnica de coleta para busca federada usando o protocolo *Open Archives Initiative - Protocol for Metadata Harvesting (OAI-PMH)* ou, ainda, outros protocolos (HIGGINS, 2007a).

A próxima seção contextualizará os desafios que a gestão de dados de pesquisa apresentam, considerando que os dados são altamente heterogêneos e não estruturados, na cauda longa da ciência.

2.10. AS BIBLIOTECAS E A GESTÃO DE DADOS DE PESQUISA

A gestão de dados de pesquisa representa possibilidade de transformação no papel das bibliotecas, uma vez que pode impactar suas operações, funções e, possivelmente, até sua missão (ARL, 2007). O paradigma da *eScience*, à medida em que é pautado pela geração, uso e compartilhamento intensivo de dados (SAYÃO; SALES, 2017a), afeta todas as áreas, inclusive as pesquisas realizadas na chamada cauda longa da ciência. Anuncia, de acordo com Hey e Hey (2006), um novo e excitante mundo para as bibliotecas, que serão colocadas no centro do desenvolvimento da infraestrutura de pesquisa da próxima geração. Conforme Gold (2007a), as atividades de pesquisa, nas últimas décadas, geraram quantidades enormes de dados, levando ao desafio de desenvolver práticas para gerenciar e fornecer acesso a esses dados.

Segundo o relatório final da *Joint Task Force on Library Support for e-Science* da ARL, havia sensação crescente de que as tendências da *eScience* estavam evoluindo rapidamente e que as bibliotecas estavam perdendo oportunidades de contribuição e engajamento, à medida que essa forma de pesquisa evoluía. Para Corral, Kennan e Afzal (2013), a participação da biblioteca na gestão de dados de pesquisa é recente. Estudos de caso sobre o papel das bibliotecas como gestoras de dados começaram a surgir em 2008. Contudo, Cox *et al.* (2017) afirmam que o potencial da biblioteca acadêmica na gestão de dados de pesquisa tem sido discutido há mais de uma década, embora a natureza e a extensão do protagonismo delas permaneçam pouco claras.

De acordo com Frederick (2016), as bibliotecas têm interesse nos dados, pois seus usuários precisam recuperar e reutilizar dados de pesquisa. Os usuários buscam as bibliotecas para obter assistência. Já os governos e organizações não governamentais enxergam as bibliotecas como local onde podem disponibilizar seus dados para análise e reutilização.

Defendendo as bibliotecas como gestoras de dados de pesquisa, Henty (2008) argumenta que elas se encontram bem envolvidas no ciclo de comunicação acadêmica, por meio da organização e fornecimento de informações (publicações, manuscritos, audiovisuais e outros formatos) e, mais recentemente, no oferecimento de repositórios institucionais. Até agora, as bibliotecas estiveram mais envolvidas com o ensino, situação que pode mudar a partir do momento em que surge a necessidade de ela se envolver mais com a pesquisa. Assim, há mudança de papéis, alavancada pela crescente disponibilização de dados primários originados a partir das pesquisas científicas.

Lewis (2010), contudo, propõe avaliar se a gestão dos dados de pesquisa seja realmente trabalho pertinente às bibliotecas universitárias. O autor afirma que a resposta é ao mesmo tempo, um sim e um não de forma direta. ‘Sim’, porque os dados de pesquisa representam parte da base de conhecimento global, uma extensão natural do papel atual da biblioteca universitária. Por outro lado, a resposta é ‘não’ porque os desafios em termos de infraestrutura, habilidades e mudanças culturais requerem ação orquestrada por várias partes interessadas e não apenas as bibliotecas universitárias.

Lewis (2010) adverte que fazer essa avaliação não significa má vontade com a questão da gestão de dados por parte das bibliotecas universitárias, que exercem papel fundamental no desenvolvimento da capacidade do setor de educação superior, no sentido de gerenciar os ativos de dados de pesquisa. Segundo o autor, medidas precisam ser tomadas para a compreensão do contexto, mesmo os recursos para fazer progressos imediatos sejam localmente escassos.

Para aprofundar a discussão, a seção seguinte abordará a Biblioteconomia de Dados, uma área emergente diante dos novos papéis e desafios postos para a área de Biblioteconomia e Ciência da informação.

2.10.1. Biblioteconomia de dados

Os serviços oferecidos pelas bibliotecas expandiram-se – ou ainda precisam se expandir – para além dos serviços tradicionais de forma a incluir atividades em torno dos dados de pesquisa (QIN, 2012). Em meio às mudanças nos papéis das bibliotecas e bibliotecários e o surgimento dos serviços de dados em bibliotecas, Gold (2007b) traz nova visão para a

chamada Biblioteconomia de Dados (*data librarianship*), também chamada por Qin (2012) de *eScience librarianship*.

A Biblioteconomia de Dados é

[...] uma mistura evolutiva de habilidades tradicionais da Biblioteconomia, como o aconselhamento sobre a descoberta de recursos, por um lado, o fornecimento de informações sobre questões específicas de formatos de dados ou obsolescência digital, por outro (RICE; SOUTHALL, 2016, p. 16, tradução nossa).

Assim, a Biblioteconomia de Dados pode, em algumas ocasiões,

[...] exigir a aplicação de conhecimentos especializados detalhados sobre coleções ou análise de empreendimentos. Em outras ocasiões, pode exigir o fornecimento de conselhos sobre direitos autorais, gestão de dados como parte do processo de pesquisa, ou seleção das ferramentas de análise de dados mais apropriadas (RICE; SOUTHALL, 2016, p. 16, tradução nossa).

Para Gold (2007b), a Biblioteconomia de Dados desenvolve-se a partir dos serviços especializados em dados, desempenhando um papel mais amplo nas fases de pré-publicação (*upstream*), quando os dados estão sendo gerados e analisados (início do ciclo de pesquisa), e quando na pós-publicação (*downstream*). O ciclo de vida dos dados começa no início do processo científico. Os serviços das bibliotecas desenvolvem-se em torno das etapas iniciais de pesquisa, em parceria com os pesquisadores, mas, também, na pós-publicação, como apoio à publicação de dados, criação de conjunto de dados, elaboração de manuais, tutoriais, inclusão de metadados, preservação *etc.*

A Biblioteconomia de Dados avançaria se aproveitasse o já demonstrado interesse dos bibliotecários nos sistemas de publicação da ciência. Com a crescente disponibilização de dados, as editoras comerciais estão interessadas tanto na criação de ferramentas para gestão e descoberta de dados quanto na publicação de conjuntos de dados em revistas de dados revisadas por pares, expandindo, assim, a noção tradicional de publicação. Tal cenário exige dos bibliotecários novas habilidades na inclusão e gerenciamento dos conjuntos de dados e, também, na criação de padrões relacionados à publicação, descrição, citação, descoberta, e reutilização de dados e, ainda, políticas para divulgação (GOLD, 2007b).

A Biblioteconomia de Dados não pode significar apenas mudança da organização de informações para a organização de dados, mas uma nova postura em que bibliotecas e bibliotecários se incorporem no processo de pesquisa (SWANSON; RINEHART, 2016). Gold (2007b, *online*, tradução nossa) adverte que “se quiserem desempenhar novos papéis na Biblioteconomia de Dados, as bibliotecas precisarão de recursos adequados [...]” tanto humanos quanto financeiros. Martinez e Macdonald (2005) destaca o reconhecimento

profissional para a área, sugerindo introduzir a Biblioteconomia de Dados como módulo ou componente das qualificações nas áreas de Biblioteconomia e Ciência da informação. Além disso, poderia oferecer uma certificação em Biblioteconomia de Dados por meio das organizações profissionais.

Em meio a uma Biblioteconomia que se desenvolve em torno dos dados (GOLD, 2007b; QIN, 2012), mantendo as habilidades e papéis tradicionais (RICE; SOUTHALL, 2016), surgem novos desafios e demandas para as bibliotecas. Isso leva à consequente alteração ou adaptação dos papéis dessas instituições e de seus profissionais, questões abordadas nas seções seguintes.

2.10.2. Bibliotecas: ampliação de papéis

As bibliotecas podem e devem assumir vários papéis em relação aos dados de pesquisa, trabalhando em parcerias - serviço de tecnologia da informação (TI), departamentos de pesquisa e gestores. Martinez e Macdonald (2005) apontam que as bibliotecas podem atuar oferecendo repositório de dados, desenvolvendo e preservando coleções locais; fornecendo serviços de referência de forma a ajudar pesquisadores a identificar recursos adequados e soluções para problemas. Podem educar usuários a liderem com acesso a dados e recursos por meio de atividades de ensino e aprendizagem.

Lewis (2010) discorre sobre nove áreas onde será necessário grande engajamento da biblioteca, sejam elas estratégicas, táticas ou operacionais: 1) conscientização da equipe; 2) aconselhamento; 3) conscientização do pesquisador; 4) alfabetização em dados; 5) integração dos estudantes de graduação; 6) desenvolvimento de infraestrutura local; 7) identificação de habilidades; 8) condução de políticas; e 9) influência na política nacional de dados.

O desenvolvimento e conscientização da equipe tem como meta aumentar o nível geral de conscientização da equipe da biblioteca em relação à *eScience* e questões de gestão de dados, com o objetivo de equipar o pessoal para realizar conversas com colegas acadêmicos e estudantes de pesquisa sobre esses tópicos.

À medida que a equipe melhora os conhecimentos sobre a gestão de dados, a biblioteca deve fornecer aconselhamento sobre gestão de dados para pesquisadores, tanto informalmente quanto através do desenvolvimento de conteúdo formal em *sites* de bibliotecas.

Paralelamente ao fornecimento de aconselhamento a equipes ou pesquisadores individuais, as bibliotecas universitárias devem conscientizar seus pesquisadores, discutindo os desafios da gestão de dados em suas instituições, por meio de diferentes canais disponíveis.

Além disso, as bibliotecas devem ensinar/desenvolver alfabetização em dados (*data literacy*) nos estudantes de pós-graduação, bem como a todos os pesquisadores. A *data literacy* seria uma evolução da formação em competência informacional (*information literacy*) já desenvolvido pelas bibliotecas. A alfabetização em relação à gestão de dados de pesquisa é fundamental, pois, representa uma oportunidade de influenciar a maneira como futuros pesquisadores abordam os dados no planejamento de suas pesquisas.

As bibliotecas devem também incluir o conteúdo de gestão de dados na graduação, levando a importância/uso de dados para os estudantes, estendendo o desenvolvimento de habilidades de gestão de dados para estudantes de graduação que começarão a desenvolver boas práticas e levar o uso de dados de pesquisa reais para melhorar a experiência de aprendizagem desses alunos.

Outra importante área de engajamento é o desenvolvimento de capacidade local (infraestrutura) para armazenamento e curadoria de dados. A criação de repositórios institucionais seria exemplo de infraestrutura, ponto de partida para a curadoria de dados.

A identificação das habilidades e competências relacionadas à gestão de dados é outra área de atuação das bibliotecas. Elas devem trabalhar junto às escolas de Biblioteconomia e Ciência da informação, identificando as lacunas de habilidades para desenvolver novos recursos de treinamento para supri-las.

Exercer liderança na condução da política local de dados também é uma área de atuação das bibliotecas. Os contatos informais com pesquisadores e outras partes interessadas da pesquisa representam oportunidade para que elas exerçam grau de liderança política de mais formal em nível universitário. Os comitês de pesquisa e inovação das universidades e as equipes de alta administração precisam entender a natureza do desafio da gestão de dados e os benefícios de uma abordagem de dados coerente em toda a instituição.

Outra área de engajamento seria a atuação em nível macro, influenciando na política nacional de dados. Os bibliotecários devem participar do desenvolvimento das políticas nacionais para dados de pesquisa.

As áreas de engajamento foram agrupadas em uma pirâmide (FIG. 9).

FIGURA 9 - Áreas de engajamento das bibliotecas na gestão de dados de pesquisa

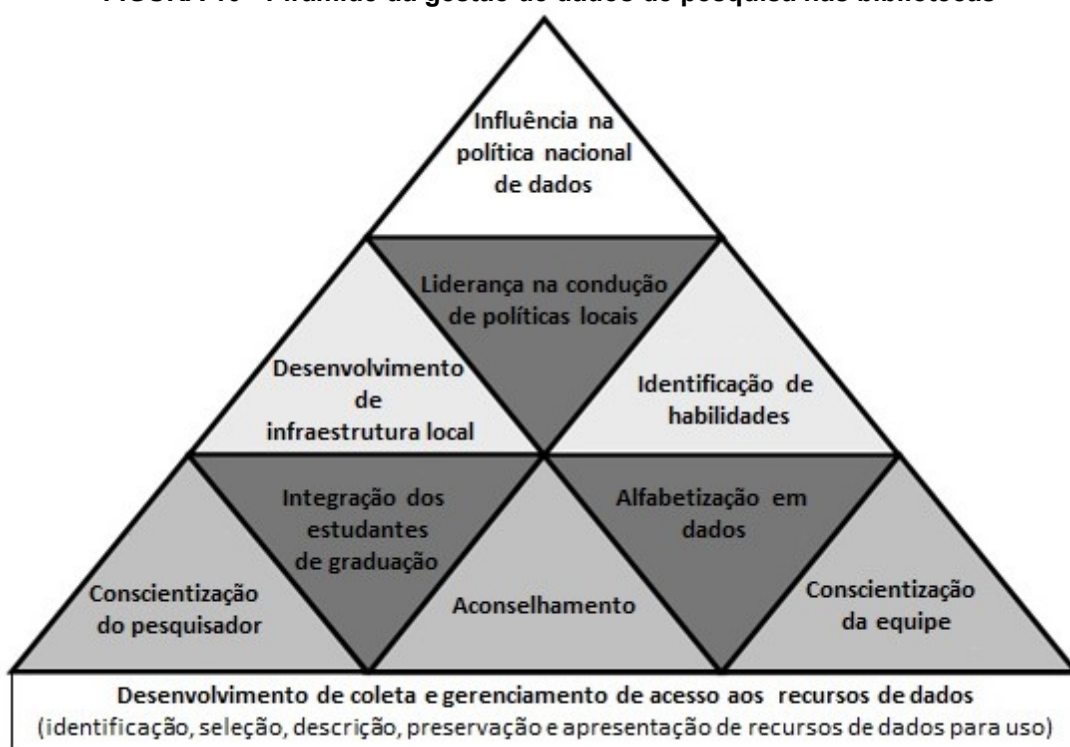


Fonte: (LEWIS, 2010, p. 157).

Lewis (2010) identifica que as atividades mais baixas na pirâmide são áreas de engajamento precoce e podem ser apropriadas para o maior número de bibliotecas universitárias, independente da instituição que estejam ligadas. Portanto, seriam questões mais fáceis de serem realizadas.

Corrall (2012) acrescenta à base da pirâmide de Lewis (2010) o desenvolvimento da coleta de dados, atividade discutida por Newton, Miller e Bracke (2010). Assim, a pirâmide de gestão de dados de pesquisa nas bibliotecas (FIG. 10), apesar de não contemplar o ciclo de vida completo dos dados, tem como base o desenvolvimento de coleta e gerenciamento de recursos de dados (processos de identificação, seleção, descrição, preservação e apresentação de recursos de dados para uso), processos em que as bibliotecas já possuem muita experiência.

FIGURA 10 - Pirâmide da gestão de dados de pesquisa nas bibliotecas



Fonte: (CORRALL, 2012, p. 111, tradução nossa).

Tenopir, Birch e Allard (2012) apontam algumas ações que as bibliotecas deverão tomar para apoio à gestão de dados de pesquisa nas instituições a que estão ligadas: 1) reconhecer que os pesquisadores precisam de serviços de dados de pesquisa e que a biblioteca pode desempenhar esse papel; 2) determinar o curso que a biblioteca irá tomar – criando novo perfil em sua comunidade com o oferecimento de serviços de dados de pesquisa, expandindo o papel da biblioteca ou fortalecendo o perfil existente com a expansão dos serviços tradicionais; 3) construir um conjunto de serviços que faça sentido para a comunidade no qual está vinculada, levando em consideração a variedade de serviços de dados de existentes; 4) identificar as outras disciplinas em que o serviço de dados pode oferecer suporte à pesquisa, além das ciências; 5) apoiar o desenvolvimento profissional do corpo docente e da equipe da biblioteca para a obtenção do conhecimento dos serviços de dados de pesquisa; 6) considerar a criação de cargo/função de bibliotecário de dados (*data librarian*) para liderar a iniciativa de serviços de dados de pesquisa; e 7) conectar-se com os departamentos de pesquisa no *campus* para colaborar, colocando a biblioteca como o centro do serviço de dados de pesquisa.

Kim (2013), em artigo falando especificamente sobre o compartilhamento de dados no contexto da gestão de dados científicos, aponta os seguintes desafios para as bibliotecas:

- a) aumentar a conscientização dos pesquisadores sobre compartilhamento de

dados e práticas de reutilização - por meio de aconselhamento informal, consultoria, classe e/ou desenvolvimento profissional, fornecimento de guias e tutoriais práticos e técnicos sobre gestão de dados e compartilhamento, desenvolvimento de *sites* com orientação sobre gestão de dados (documentação de dados, métodos para compartilhamento, confidencialidade e propriedade intelectual dos dados, dentre outros); promoção de *workshops* e/ou cursos de gestão de dados, além de alguns canais menos formais, como consultorias ou treinamentos individuais;

- b) fornecer serviços de arquivamento de dados – as bibliotecas/instituições podem criar ou expandir seus repositórios institucionais (preferencialmente disciplinares) para permitir também o arquivamento dos dados de pesquisa, com especial atenção às políticas que precisam abordar questões complexas relacionadas ao acesso e suporte – acordos de licenciamento para reuso com diretrizes claras, períodos de embargo quando acordado com o pesquisador. Algumas instituições ainda oferecem plataformas de conjunto de serviços que facilita ou aprimora o processo de pesquisa colaborativa, o compartilhamento e a publicação de dados de pesquisa;
- c) construir mão de obra para a chamada Biblioteconomia de Dados – as bibliotecas devem ser desafiadas a desenvolver uma nova vertente profissional - a Biblioteconomia de Dados – treinando novos bibliotecários e capacitando aqueles que já estão trabalhando na área de dados.

De acordo com Cox e Pinfield (2014), o papel na gestão de dados de pesquisa é um dentre os papéis que as bibliotecas podem oferecer no suporte a pesquisa. Cox e Pinfield (2014) apontam outros papéis no apoio à pesquisa, compilados a partir de Auckland (2012), Brewerton (2011) e Garritano e Carlson (2009): 1) aconselhamento sobre fontes de financiamento; 2) papéis embutidos ou de suporte que realizam revisões de literatura ou alertas de conscientização atuais para projetos ou grupos de pesquisa; 3) bibliometria e métricas de impacto; 4) apoio ao *Research Excellence Framework*²⁶; 5) treinamento em *software* bibliográfico; 6) *advocacy*²⁷ para acesso aberto/repositório institucional; 7) aconselhamento de análise de dados; 8) aconselhamento sobre questões de direitos autorais; e 9) aconselhamento sobre arquivamento de registros de pesquisa.

²⁶ Processo de revisão de especialistas usado para avaliar a qualidade da pesquisa e seus resultados, o impacto mais amplo da pesquisa e a vitalidade do ambiente de pesquisa em instituições de ensino superior (AUCKLAND, 2012).

²⁷ *Advocacy* – “[...] defender e argumentar em favor de uma causa, uma demanda ou uma posição [...] iniciativas de incidência ou pressão política, de promoção e defesa de uma causa e/ou interesse, e de articulações mobilizadas por organizações da sociedade civil com o objetivo de dar maior visibilidade a determinadas temáticas [...]” (LIBARDONI, 2000, p. 208).

Já Gabridge (2009) discorre sobre os papéis das bibliotecas em relação à gestão de dados: 1) análise dos requisitos de depósito do conjunto de dados; 2) planejamento da gestão de dados; 3) ensino de boas práticas de dados aos alunos; 4) coleta e disseminação dos conjuntos de dados; 5) definição padrões para preservação de dados.

Pela posição dos diversos autores, nota-se que há semelhança entre as funções no suporte à gestão de dados de pesquisa propostas. Há certa uma convergência de ideias em torno do papel que as bibliotecas devem desempenhar no suporte à gestão de dados de pesquisa. Na próxima seção serão apresentados os papéis e habilidades que os bibliotecários devem possuir ou desenvolver para desempenhar as atividades voltadas para a gestão de dados.

2.10.3. Os bibliotecários e a gestão de dados de pesquisa: papéis e habilidades

Os bibliotecários são peças-chave, pois já desenvolvem atividades como a organização e gestão de informações, extensíveis à gestão dos dados de pesquisa (COX; PINFIELD, 2014).

Latham (2017) lembra que os bibliotecários sempre precisaram reinventar seus papéis, principalmente no último quarto de século, expandindo suas experiências e competências tecnológicas. Mais recentemente, as bibliotecas tendem a assumir a gestão de dados de pesquisa, já que os dados parte essencial do registro acadêmico (TENOPIR *et al.*, 2017).

O papel desses profissionais na gestão dos dados não é estático, imutável. Varia dependendo da necessidade e apoio institucional. Assim,

[...] algumas bibliotecas estabeleceram colaborações nessas áreas e se integraram nas atividades de gestão de dados, outras bibliotecas estão no início das etapas de ajudar os pesquisadores com seus desafios de gestão de dados (REED, 2015, p. 2, tradução nossa).

Em algumas bibliotecas, os bibliotecários já vivenciam mudanças, ao assumirem novas responsabilidades. Em outras, novos postos especializados estão sendo criados para atender à demanda (AUCKLAND, 2012).

Macdonald e Martinez-Urbe (2013) alegam que o gestor não configura emprego novo. No entanto, pressões políticas e o movimento dos dados abertos nos últimos anos levaram os bibliotecários ao centro das atenções – sendo chamados de bibliotecários de dados. O papel do bibliotecário de dados é fluído. Pode até parecer que seja político, de estratégia e *advocacy*, aliado aos trabalhos técnicos (preservação, inclusão de metadados e orientação), mas os autores entendem que tudo se trata basicamente de ‘apoiar a pesquisa’.

São várias as áreas onde os bibliotecários desempenham papéis. Podem: atuar consultoria e escrita de planos de gestão de dados para submissão às agências de financiamento; ajudar na determinação dos padrões de metadados; na curadoria e arquivamento de dados; e na descoberta e citação de repositórios ou conjunto de dados (REED, 2015).

Cox e Pinfield (2014) destacam os papéis em que os bibliotecários largas experiências possuem e que poderão ser aproveitadas na gestão dos dados de pesquisa. São elas: na determinação dos padrões de metadados – importantes na recuperação e curadoria de dados; na assessoria e treinamentos - uma ampliação dos serviços de treinamento em competência informacional de forma a abarcar a gestão de dados; e nas amplas redes de relacionamentos com os departamentos acadêmicos que podem facilitar práticas de compartilhamento.

Gold (2007b) acrescenta que a experiência de domínio específico do conhecimento é essencial para a gestão de dados de pesquisa. Os bibliotecários precisarão estabelecer relações com equipes de pesquisa, seja para treinar os pesquisadores (peritos do domínio) sobre gestão e curadoria de dados ou sobre as necessidades de infraestrutura e serviços de um domínio. A autora defende que criar certa fluência nas atividades das bibliotecas (gestão e curadoria) ou na cultura científica dos grupos de pesquisa será essencial para a nova área denominada 'Biblioteconomia de Dados' (GOLD, 2007b).

Gold (2007b) ressalta os papéis dos bibliotecários em dois estágios presentes no ciclo de vida dos dados de pesquisa: 1) pré-publicação (*upstream*); e 2) pós-publicação (*downstream*).

No estágio pré-publicação (*upstream*) (1) destacam-se: a colaboração no processo de pesquisa, que significa a criação de protótipos de curadoria de dados ou apoio no uso de documentação, práticas ou padrões que garantam a longevidade dos dados; e a criação de repositórios mais dinâmicos que suportem fluxos de trabalho pré-publicação, incluindo ambientes de colaboração que suportem integração, análise e visualização de dados.

Já no estágio pós-publicação (*downstream*) (2) destacam-se: a seleção, aquisição e licenciamento de dados de conjunto de dados; a criação de metadados (ou padrões de metadados) para descoberta e descrição de conjuntos de dados; a criação ou organização de documentação relacionada a dados – elaboração de manuais, tutoriais para apoio à publicação de dados; a oferta de serviços de preservação para dados digitais; a assessoria na avaliação e seleção dos dados a serem mantidos em longo prazo; a ajuda aos usuários a encontrar dados relevantes para sua pesquisa; o trabalho junto a sociedades acadêmicas e editoras acadêmicas no desenvolvimento de padrões e sistemas de publicação de dados -

fluxos de trabalho de publicação, esquemas de identificadores globais, esquemas de vinculação (*linking schemes*), padrões de limpeza e normalização de dados, padrões para fornecimento de crédito e reconhecimento aos autores, propriedade intelectual; e a criação de repositórios de dados.

Martin (2016) propõe um quadro com uma estrutura de gestão de dados centrada no usuário (FIG. 11). Nele, os usuários são o centro do processo, não os dados. O quadro destaca o papel do bibliotecário na gestão de dados – chamada pela autora de ciência de dados (*data science*). Os papéis incluem: serviços; melhores práticas para a gestão de dados; *literacy* (alfabetização); arquivamento e preservação; e políticas.

FIGURA 11 - Estrutura de gestão de dados centrada no usuário



Fonte: (MARTIN, 2016, p. 2, tradução nossa).

Nessa estrutura, os serviços incluem atividades como: avaliar as necessidades dos pesquisadores; realizar verificação ambiental dos dados institucionais; conduzir entrevistas de referência; projetar conjunto de serviços, inclusive planos de gestão de dados com base nas necessidades dos usuários *etc.* As melhores práticas de gestão de dados abrangem: a recomendação de convenções de nomenclatura de arquivos; o auxílio no uso de caderno eletrônico de laboratório (*e-notebook*)²⁸ e outras atividades. A alfabetização em dados

²⁸ E-notebook - programa de computador usado para substituir os cadernos de laboratório em papel, usados para documentar pesquisas, experiências e procedimentos usados em laboratórios. Usados como documento legal em questões de prova (WIKIPEDIA, 2017).

(*literacy*) pode oferecer aulas pessoalmente ou *online* de tópicos como: práticas recomendadas de gestão de dados; ferramentas de dados (*R*²⁹ e *Python*³⁰); e *software* de visualização de dados (*Tableau*³¹). O arquivamento/preservação foca nos repositórios de dados - institucional e de assunto - e suas respectivas opções de preservação (LOCKSS³², CLOCKSS³³). No campo das políticas, é aconselhável trabalhar em colaboração com organismos de padronização local, institucional, regional, nacional e internacional a fim de criar padrões de compartilhamento de dados (MARTIN, 2016).

Nesse sentido, Macdonald e Martinez-Urbe (2013) destacam uma série de atividades que os profissionais - bibliotecários de dados - podem assumir para apoiar os pesquisadores na busca e uso dos recursos de dados. São elas: treinamentos; entrevistas de referência; preparação e compilação de guias de usuário; catalogação de recursos; correspondência dos dados gerados pelos pesquisadores com as fontes de dados reconhecidas; solução de problemas relacionados a dados; ensino; interpretação do *codebook*/questionário/documentação; gestão de dados; capacitação/conscientização por meio de listas de discussão; conferências, seminários e *workshops*; e participação como representante institucional nos centros de dados nacionais.

Em relação às habilidades que o profissional deve apresentar, Corral (2012) corrobora Gold (2007b) e ressalta a necessidade de conhecimento de domínio e habilidades técnicas em gestão de dados e áreas de competência específicas como o aspecto cultural de cada disciplina, metodologias de pesquisa, fluxos de trabalho científicos, gestão do ciclo de vida, pensamento sistêmico e conhecimento de tecnologia. Martinez e Macdonald (2005) apontam que os profissionais devem desenvolver habilidades para descoberta de dados, criar ferramentas para tornar os dados utilizáveis e familiaridade na construção do conjunto de dados, habilidades para facilitar o acesso, e oferecer suporte e treinamento aos que desejam usar dados de pesquisa.

Garritano e Carlson (2009) identificaram cinco categorias de conjuntos de habilidades necessárias ao bibliotecário no apoio à gestão de dados de pesquisa: 1) experiência em bibliotecas e ciência da informação; 2) conhecimento profundo na área disciplinar da

²⁹ R – *software* livre para computação estatística e gráficos (R FOUNDATION, 2017).

³⁰ Python – linguagem de programação (PYTHON SOFTWARE FOUNDATION, 2017).

³¹ Tableau – *software* proprietário para visualização de dados (TABLEAU SOFTWARE, 2017).

³² LOCKSS (*Lots of Copies Keep Stuff Safe*) - sistema de preservação digital de código aberto baseado no princípio de que "muitas cópias mantêm as coisas seguras". As bibliotecas usam o LOCKSS para criar e preservar suas próprias coleções digitais locais, tomando propriedade do conteúdo, criando coleções locais permanentes (STANFORD UNIVERSITY, 2017).

³³ CLOCKSS (*Controlled LOCKSS*) – arquivo global, que usa a tecnologia LOCKSS para preservar o conteúdo em 12 locais diferentes espalhados pelo mundo visando otimizar a segurança do conteúdo contra ameaças políticas e ambientais (CLOCKSS, 2017).

pesquisa; 3) parcerias e divulgação (internas e externas); 4) participação em pesquisas; 5) equilíbrio da carga de trabalho.

Carlson (2006) aponta que a gestão dos dados não deve ficar por conta dos próprios pesquisadores, mas de bibliotecários. Esses terão que definir, categorizar e arquivar o volume e o fluxo de dados gerados nos experimentos.

Corrêa (2017) reconhece que a maioria das diretrizes para gestão de dados é destinada aos pesquisadores. Porém, os bibliotecários têm papel cada vez mais importante. No entanto, há poucos estudos sobre como estimular as competências necessárias a esse profissional para a gestão de dados. Corral (2012) arremata dizendo que as funções e responsabilidades na gestão de dados de pesquisa não estão resolvidas, embora haja oportunidade real para os bibliotecários assumirem liderança dentro de suas instituições e da comunidade de pesquisa.

Não há evidência na literatura de participação de bibliotecários junto aos pesquisadores na geração de dados. Porém, urge que esses profissionais tenham conhecimento dos conjuntos de dados existentes, assessorando e apoiando pesquisadores na identificação de dados pré-existentes úteis e relevantes para suas pesquisas (AUCKLAND, 2012).

Nesse sentido, Swanson e Rinehart (2016, p. 98, tradução nossa) acreditam que a mudança no papel do bibliotecário e o surgimento dos serviços de dados em bibliotecas “[...] decorrem da necessidade de se envolver com pesquisadores ao longo do processo de pesquisa, incluindo os estágios iniciais do processo, quando os resultados estão em sua infância” e não do surgimento dos dados em si.

Assim como na seção anterior, percebe-se que há consonância de pensamento entre os autores no que se refere às habilidades e papéis que os bibliotecários devem desenvolver no suporte à gestão de dados. Para que as bibliotecas consigam se adaptar a essas novas demandas é necessário assumir novos papéis e desenvolver habilidades, como a de compreender os pesquisadores, suas práticas e necessidades em relação aos dados, os tipos de dados gerados, as distinções disciplinares *etc.* A seção seguinte aborda de forma breve algumas dessas questões.

2.11. PRÁTICAS E NECESSIDADES DOS PESQUISADORES

Todos os atores envolvidos na infraestrutura de pesquisa como pesquisadores, pessoal de TI e bibliotecas precisam enfrentar o desafio de gerenciar e compartilhar pesquisas e dados para permitir novos conhecimentos científicos e impulsionar a inovação (GABRIDGE, 2009). Contudo, há grande incerteza em relação aos papéis que as bibliotecas devem

desempenhar, pois as infraestruturas, as práticas e requisitos de domínio e os trabalhos de pesquisa variam amplamente. (GOLD, 2007b).

Dessa forma, torna-se crucial aos bibliotecários o conhecimento das distinções entre diferentes tipos de pesquisadores. Devem compreender as atividades que realizam durante o ciclo de vida da pesquisa. É aí que incidem muitos serviços das bibliotecas e onde potenciais novos serviços podem ser identificados (AUCKLAND, 2012). Logo, um dos desafios das bibliotecas universitárias é atender as necessidades de “[...] uma clientela diversificada” (REED, 2015, p. 2, tradução nossa), com necessidades também variadas.

Para que os serviços oferecidos pelas bibliotecas consigam atender às demandas, Gabridge (2009, p. 15, tradução nossa) destaca que é importante conhecer a necessidade dos investigadores. Sendo assim, “estudar as necessidades de pesquisadores individuais e da instituição como um todo é componente importante do trabalho [...] da biblioteca”.

Às bibliotecas e aos bibliotecários cabem estudar as necessidades de pesquisadores e da instituição como um todo. Por outro lado, os pesquisadores também devem aprender a enxergar que os bibliotecários são profissionais capazes de ajudá-los a resolver seus problemas de dados, pois as bibliotecas possuem longa experiência na gestão e acesso à informação (GABRIDGE, 2009).

O reconhecimento da diferença entre as necessidades de informação dos pesquisadores não é novo. Tenopir *et al.* (2015, p. 3, tradução nossa) afirmam que “[...] os bibliotecários acadêmicos buscam entender as necessidades dos pesquisadores em suas instituições e desenvolver estratégias específicas para atender a essas necessidades”.

Auckland (2012) identificou que os pesquisadores não são grupo homogêneo. Eles têm atividades, discursos, abordagens e necessidades diferenciadas, principalmente, em relação à disciplina ou assunto pesquisado. Sua cultura, práxis e estágio de carreira diferem individualmente.

Latham (2017) atenta para essas questões ao afirmar que a criação e a gestão de dados variam muito entre os próprios pesquisadores ou entre os domínios de pesquisa. A criação e a gestão de dados nas ciências duras acontece de forma distinta das humanidades. Os bibliotecários devem considerar essas diferenças, conhecendo as perspectivas de pesquisa de todas as disciplinas. Além disso, segundo Weller e Monroe-Gulick (2015), à medida que a pesquisa evolui, as práticas e necessidades também evoluem. Será necessário para bibliotecas e instituições acadêmicas em geral continuarem a monitorar as práticas de pesquisa de dados e necessidades dos pesquisadores em todas as áreas do conhecimento.

Uma série de estudos examinou as práticas de dados dos professores e estudantes de pós-graduação (WELLER; MONROE-GULICK, 2015). Abordaremos alguns estudos nesta seção.

Wiley e Mischo (2016) entrevistaram 21 docentes da Faculdade de Engenharia e Ciência Atmosférica da *University of Illinois Urbana-Champaign* (UIUC). Visando determinar as suas práticas e necessidades de gestão de dados no contexto de atividades de pesquisa. As entrevistas abordaram questões como: exigência de planos de gestão de dados pelas fontes de financiamento; tipos de dados; formato; descrição; uso de repositórios disciplinares e institucionais; compartilhamento de dados; mecanismos de descoberta; papel dos dados no processo de comunicação acadêmica; conscientização sobre a gestão de dados e serviços de preservação; desafios da gestão de dados de pesquisa; perspectivas sobre quem estava interessado em seus dados; vida útil dos dados; e identificação da importância dos dados para preservação ao longo do tempo. Além de examinar as práticas e atitudes em relação ao compartilhamento de dados, a pesquisa buscou informar sobre a criação de serviços de dados que melhor atendessem as necessidades dos pesquisadores e alunos.

A pesquisa concluiu que os docentes conhecem os serviços de preservação e gestão de dados oferecidos pelas bibliotecas da UIUC, principalmente a partir dos eventos de divulgação oferecidos. Porém nenhum dos entrevistados contatou ou usou os serviços de preservação ou gerenciamento disponibilizados pelas bibliotecas, que possuíam serviços destinados a conscientizar os pesquisadores sobre as políticas mandatórias das agências de financiamento, a auxiliar na análise do plano de gestão de dados e a fornecer opções de armazenamento (WILEY; MISCHO, 2016).

Dentre outros resultados, a pesquisa apontou que os pesquisadores precisam de assistência no armazenamento, *backup*, preservação em longo prazo, arquivamento de dados e soluções para os volumes de dados produzidos (WILEY; MISCHO, 2016).

Em 2013, *survey* realizada com estudantes de pós-graduação e professores da *Kansas University* (KU) buscou compreender as práticas de dados, desafios e futuras necessidades desses grupos de forma comparativa (WELLER; MONROE-GULICK, 2015).

As bibliotecas até agora concentravam seus serviços de dados ajudando no desenvolvimento de planos de gestão de dados. Porém, os resultados da pesquisa indicam que esse é o serviço com o qual os participantes têm menos problema. Apesar disso, as oficinas sobre a elaboração de planos de gestão de dados seria boa alternativa para os serviços das bibliotecas. Os alunos precisam de assistência na análise dos dados. As bibliotecas poderiam oferecer oficinas em torno de ferramentas de dados específicos ou técnicas analíticas, firmando em parcerias com departamentos ou centro de pesquisa. Há

ainda necessidade de encontrar opções de armazenamento de dados seguras e fáceis de usar. Dessa forma, poderiam ser realizadas oficinas de gestão de dados por bibliotecários ou por outros departamentos do *campus* (WELLER; MONROE-GULICK, 2015).

McLure *et al.* (2014), em estudo realizado em 2012 na *Colorado State University (CSU)* com 31 participantes - professores, pesquisadores e pesquisadores associados -, utilizando a técnica de grupo focal, buscou identificar a natureza dos conjuntos de dados que os pesquisadores criam ou mantêm. Buscou também entender como os participantes gerenciam seus dados; qual a necessidades de apoio em relação ao compartilhamento, curadoria e preservação de seus dados; e qual a viabilidade de adaptação do protocolo de entrevistas *Data Curation Profiles (DCP) Toolkit* desenvolvido pela *Purdue University Libraries* para uso em grupos focais com pesquisadores.

Dentre outros resultados, a investigação produziu descrições detalhadas de diversos projetos de pesquisa; prevalência da criação ou coleta em oposição à reutilização de dados criados ou coletados por outros; indicação de formatos de arquivos de dados específicos comumente usados; prevalência e predominância de um grande número de pequenos (200 GB ou menos) arquivos de dados em muitos projetos de pesquisa. Além disso, os participantes indicaram que gostariam de assistência na conservação e compartilhamento de dados. Discorreram sobre seus déficits de habilidades e conhecimentos. Os participantes demonstraram que não se importam com qual instituição deverá fornecer a assistência – seja biblioteca ou outro local dentro da universidade –, desde que sejam adequadamente atendidas suas necessidades. Os pesquisadores demonstraram interesse principalmente em: treinamentos com foco na coleta digital de dados em substituição ao uso de cadernos de laboratório de papel; gestão de dados; novas metodologias para gravação; e abordagens e ferramentas de organização de dados (MCLURE *et al.*, 2014).

Em estudo de caso realizado em 2013 na *Oregon State University (OSU)* com professores, pesquisadores e pós-doutorandos, os autores procuraram demonstrar como o conhecimento local das práticas de gestão de dados pode ajudar no desenvolvimento progressivo dos serviços de dados de pesquisa oferecidos pelas bibliotecas, onde já é oferecido suporte. Assim, a pesquisa objetivou melhorar a compreensão das práticas, opiniões, preocupações e necessidades em relação aos dados na OSU. Para esse estudo foi realizado *survey* com perguntas abordando as principais questões da gestão de dados: políticas; funções e responsabilidades; características de dados gerados; práticas de gestão de curto prazo; serviços de gestão de dados e suporte; financiamento de gestão de dados; padrões para dados de pesquisa e documentação; compartilhamento de dados; preservação em longo

prazo; e identificação de tarefas da gestão em que os pesquisadores apresentam maiores dificuldades (WHITMIRE; BOOCK; SUTTON, 2015).

A pesquisa identificou que a maioria dos dados gerados é do tipo quantitativo. Quanto à quantidade, a maioria informou que gera menos de 100 GB na maioria dos projetos. Esses dados são armazenados geralmente em computadores pessoais e dispositivos de armazenamento externo. O armazenamento em servidor da universidade é pouco usado. Além disso, a pesquisa demonstrou que os serviços de dados desenvolvidos pelas bibliotecas têm pouca visibilidade. Apenas 13% dos participantes relataram estar cientes dos serviços relacionados com o desenvolvimento ou revisão dos planos de gestão de dados oferecidos pelas bibliotecas da *Oregon State University* (OSU). Mais da metade dos participantes declararam criar seus próprios metadados a usar esquemas de metadados padrão. Isso tem implicações em questões de interoperabilidade com outros conjuntos de dados, detecção (por máquinas e humanos) e documentação (WHITMIRE; BOOCK; SUTTON, 2015).

Embora os resultados não sejam passíveis de generalizações, percebeu-se que os pesquisadores geram uma variedade de tipos de dados com pequeno volume (até 200 GB), que são armazenados em uma diversidade de locais. Os dados gerados variam muito entre as áreas do conhecimento. Os pesquisadores precisam de assistência em diversas questões relacionadas à gestão de dados de pesquisa: plano de gestão de dados; inclusão de metadados; uso de repositórios; armazenamento; preservação, dentre outros, apresentando oportunidades para a participação das bibliotecas nesse processo.

A seção seguinte define e apresenta os serviços de suporte à gestão de dados de pesquisa que podem ser realizados pelas bibliotecas acadêmicas. Além disso, aponta os caminhos para o desenvolvimento desses serviços nas instituições; classificações de serviços; e os desafios para o oferecimento de SSGDPs nas bibliotecas.

2.12. SERVIÇOS DE SUPORTE À GESTÃO DE DADOS DE PESQUISA

Há pouco tempo, o artigo científico era considerado o produto final do processo de pesquisa e o principal meio de disseminação do conhecimento científico (COATES, 2014). Contudo, as atividades de pesquisa geram grandes quantidades de dados (BARBROW; BRUSH; GOLDMAN, 2017) em formato digital, em maioria, senão em todas as disciplinas científicas (PRYOR, 2014b).

Os dados de pesquisa são, cada vez mais, vistos como produto valioso (COATES, 2014). Devem, não somente ser processados e analisados pelos pesquisadores, como gerenciados ao longo de todo o ciclo de vida (BARBROW; BRUSH; GOLDMAN, 2017). Isso cria nas

instituições a necessidade de desenvolver políticas, infraestruturas e serviços para auxiliar pesquisadores na criação, coleta, manipulação, análise, transporte, armazenamento e preservação de conjuntos de dados (PINFIELD; COX; SMITH, 2014).

Aliado ao reconhecimento do potencial e importância dos dados de pesquisa, pressões externas vêm fazendo com que a gestão de dados de pesquisa seja preocupação crescente nas universidades (ASHLEY, 2012), representando oportunidade para que as bibliotecas ofereçam serviços voltados à gestão de dados (KENNAN; CORRALL; AFZAL, 2014; KRAFT, 2015). Essas pressões referem-se ao reuso de dados e requisitos das agências financiadoras (ASHLEY, 2012); requisitos dos editores de periódicos (KNUTH; JOHNSON; HAUSER, 2015); boas práticas de pesquisa e reputação das instituições (ASHLEY, 2012); crescimento e desenvolvimento da pesquisa colaborativa; e as práticas de pesquisa aberta (*Open Access, Open Science, Open Data*) (CHIWARE; BECKER, 2018).

A gestão eficaz de dados traz recompensas ao longo e além da vida do projeto de pesquisa. É importante garantir que os dados sejam detectáveis, acessíveis e inteligíveis para permitirem a reutilização em longo prazo. Esses valores são igualmente cruciais durante a fase de pesquisa ativa. Além disso, “dados bem gerenciados levam a pesquisas de maior qualidade, maior visibilidade e aos consequentes benefícios das melhores taxas de citação” (JONES; PRYOR; WHYTE, 2013, p. 2, tradução nossa).

A gestão dos dados de pesquisa é, atualmente, tema popular na agenda acadêmica de serviços de bibliotecas e informações. Oferecem tempo propício para que as bibliotecas redefinam seu papel no apoio à pesquisa e no desenvolvimento de relacionamentos mais próximos com sua comunidade (CORRALL, 2012), constituindo oportunidade (KENNAN; CORRALL; AFZAL, 2014; KRAFT, 2015) e grande desafio para as bibliotecas (PRYOR, 2014a).

Tenopir, Birch e Allard (2012) defendem que a gestão de dados de pesquisa representa oportunidade única para que as bibliotecas acadêmicas desempenhem papel ainda mais ativo no processo de pesquisa. As maneiras são várias. As bibliotecas acadêmicas podem fornecer serviços de consultoria relacionados à gestão e conservação de dados de pesquisa. Podem também fornecer a infraestrutura para armazenamento de dados e curadoria. E, enfim, podem apoiar os bibliotecários a se tornarem membros ativos em equipes de pesquisa e de atuarem como consultores em curadoria de dados (TENOPIR; BIRCH; ALLARD, 2012).

Por outro lado, a gestão de dados de pesquisa torna-se desafio para as bibliotecas, que são vistas como grandes colaboradoras para a atividade de gestão de dados de pesquisa

(PINFIELD; COX; SMITH, 2014). Isso exige habilidades em gerenciar, pesquisar e curar grandes volumes de dados gerados (HEY; HEY, 2006). Entrar para a área do fornecimento de serviços voltados aos dados científicos ou fazer parcerias com os pesquisadores pode ser tarefa assustadora para os bibliotecários, uma vez que, segundo Monastersky (2013), a biblioteca distanciou-se do cientista. Os profissionais da área precisam se envolver ativamente com suas comunidades de usuários e entender não apenas os conceitos do domínio como também as metodologias e normas de troca acadêmica. Torna-se necessário mudar o foco da gestão de coleções para um modelo que enfatize o alcance e o engajamento dos bibliotecários no processo de pesquisa (ARL, 2007). Entretanto, apesar dos desafios, Geraci Humphrey e Jacobs (2012) defendem que os dados não são diferentes dos outros formatos de conteúdo da biblioteca. A gestão de dados representa, assim, uma extensão do papel tradicional da biblioteca no fornecimento de serviços de pesquisa e referência (TENOPIR *et al.*, 2017).

Além das bibliotecas, a gestão de dados de pesquisa representa desafio para os pesquisadores. Exige um conjunto de habilidades de gerenciamento que provavelmente não faz parte da educação do cientista. Eles provavelmente não têm tempo ou treinamento para desenvolver essas habilidades (COATES, 2014; GARRITANO; CARLSON, 2009; KNUTH; JOHNSON; HAUSER, 2015).

Tenopir, Birch e Allard (2012) afirmam que as bibliotecas acadêmicas podem desempenhar papel importante por meio das atividades desenvolvidas pelos SSGDPs, tornando-se participantes mais ativos no ciclo de criação de conhecimento dentro da instituição. Dessa maneira, segundo Tenopir *et al.* (2013, p. 77, tradução nossa) as bibliotecas estão passando por uma transição, dos serviços baseados em coleções para serviços voltados à gestão de dados de pesquisa. Isso exige “[...] redefinição de prioridades, realinhamento de responsabilidades e oferta de oportunidades para bibliotecários desenvolverem habilidades relacionadas [...] aos serviços de suporte à gestão de dados”.

Algumas bibliotecas acadêmicas começaram a oferecer serviços de gestão de dados de pesquisa às suas comunidades (BARBROW; BRUSH; GOLDMAN, 2017; KNUTH; JOHNSON; HAUSER, 2015; MULLINS, 2009). De acordo com Kennan, Corral e Afzal (2014) cresceu o envolvimento das bibliotecas com o desenvolvimento de serviços de suporte voltados para a gestão de dados. Mas há lacunas de conhecimento, habilidades e de confiança para o oferecimento de tais serviços. Os papéis e habilidades requeridos aos bibliotecários já foram tratados em seção anterior.

Em perspectiva internacional

as bibliotecas tomaram medidas importantes para estabelecer seu lugar no espaço de dados da pesquisa. Avançaram estratégica, tática e operacionalmente com práticas exemplares evidentes em áreas como formulação de políticas, desenvolvimento de repositório, inovação curricular, atualização profissional, ensino de graduação, treinamento de pós-graduação, aconselhamento de pesquisador, *advocacy* em dados, criação de perfil de curadoria e seleção de recursos (CORRALL, 2012, p. 127, tradução nossa).

No sentido de remodelar os serviços das bibliotecas para adequação às demandas da gestão de dados, Lyon (2012) propõe a *Informatics Transform* que engloba infraestruturas de *hardware*, *software* e humana.

Esta infraestrutura engloba componentes de *hardware* e *software* para integração, manipulação, recombinação e armazenamento de dados, mas também inclui a infraestrutura humana essencial necessária em nível institucional para aconselhar, orientar, treinar, coordenar e liderar o esforço de gerenciamento para desenvolver a capacidade de gestão de dados (LYON, 2012, p. 127, tradução nossa).

Lyon (2012) acredita ser agora o momento de as bibliotecas acadêmicas examinarem criticamente sua estrutura, função, portfólio de serviços, garantindo que seus serviços sejam adequados à gestão de dados de pesquisa. Essa adequação adiciona outra dimensão aos serviços oferecidos pelas bibliotecas que exigirá o “[...] uso de ferramentas inovadoras, habilidades adicionais, estruturas organizacionais alteradas, um conjunto diferente de parceiros e envolvimento em novas colaborações” (HENTY, 2008, p. 7, tradução nossa).

Embora Auckland (2012) afirme que os serviços de suporte à gestão de dados de pesquisa ainda engatinhem e que haja discussão sobre quem deve fornecê-los, as universidades e as bibliotecas estão começando a descobrir como a gestão de dados de pesquisa deve ser apoiada para qualificar os serviços de consultoria, treinamentos ou de infraestrutura para armazenamento, compartilhamento e curadoria de dados (COX; PINFIELD, 2014). Contudo, o que são e quais os serviços que as bibliotecas podem oferecer no suporte à gestão de dados de pesquisa?

Diversas nomenclaturas foram sugeridas para designar basicamente os mesmos tipos de serviços, conforme Quadro 14:

QUADRO 14 - Denominação dos serviços

Denominação	(AUTOR, ano)
<i>Data management services</i>	(BARBROW; BRUSH; GOLDMAN, 2017; CREAMER <i>et al.</i> , 2011; QIN, 2013; SCHMILLEN, 2015; SWANSON; RINEHART, 2016; WELLER; MONROE-GULICK, 2014; WHITMIRE; BOOCK; SUTTON, 2015)
<i>Data services</i>	(GARRITANO; CARLSON, 2009; GIARLO, 2013; JAHNKE; ASHER; KERALIS, 2012; LUCE, 2008; RIBEIRO; FERNANDES, 2011; SWANSON; RINEHART, 2016; WANG, 2013)
<i>Library data services</i>	(FRIEDLANDER; ADLER, 2006; WANG, 2013)
<i>Library support for data management</i>	(SWANSON; RINEHART, 2016)
<i>Library support services for research data management</i>	(LYON, 2012)
<i>Research data management services</i>	(BERMAN, 2017a; CHIWARE; BECKER, 2018; FEARON <i>et al.</i> , 2013; GOBEN; NELSON, 2018; HIOM <i>et al.</i> , 2015; JONES; PRYOR; WHYTE, 2013; KENNAN; CORRALL; AFZAL, 2014; LEE; STVILIA, 2017; MATUSIAK; SPOSITO, 2017; PRYOR; JONES; WHYTE, 2014; TENOPIR <i>et al.</i> , 2014; WHYTE; ALLARD, 2014)
<i>Research data services</i>	(KERBY, 2016; KNUTH; JOHNSON; HAUSER, 2015; SI <i>et al.</i> , 2015; TENOPIR <i>et al.</i> , 2013, 2014, 2015, 2016, 2017; TENOPIR; BIRCH; ALLARD, 2012; WILSON <i>et al.</i> , 2010; YU, H. H., 2017)
<i>Research data support services</i>	(REZNIK-ZELLEN; ADAMICK; MCGINTY, 2012)
<i>Scholarly data services</i>	(HEIDORN, 2011)
<i>Scientific data management services</i>	(OU; ZHOU, 2016)
<i>Services to support the management of research data</i>	(AUCKLAND, 2012)
Serviço de suporte à pesquisa e gerenciamento de dados	(ROSSI; BEM; LEHMKUHL, 2016)
Serviços para suporte à gestão de dados científicos	(PRÍNCIPE; SARAIVA, 2015)

Fonte: Elaborado pela autora.

Neste trabalho optamos por utilizar a designação 'serviços de suporte à gestão de dados de pesquisa' para os serviços oferecidos pelas bibliotecas no suporte a gestão de dados, por considerá-la a melhor e mais abrangente tradução possível para o conceito desses serviços.

Segundo Tenopir *et al.* (2015), os serviços de suporte à gestão de dados de pesquisa (SSGDPs) surgiram nas bibliotecas acadêmicas em resposta ao crescimento da pesquisa intensiva em dados – *eScience* - aliada às mudanças de papéis das bibliotecas e ao reconhecimento da necessidade de gestão de dados de pesquisa.

O que são esses serviços? Os serviços de suporte à gestão de dados de pesquisa (SSGDPs) são aqueles serviços oferecidos por uma biblioteca aos pesquisadores em relação à gestão de dados (TENOPIR; BIRCH; ALLARD, 2012), abordando o ciclo de vida completo dos dados, desde o planejamento da pesquisa, passando pela coleta de dados, processamento e análise, publicação e compartilhamento, preservação até a reutilização dos dados (UK DATA SERVICE, 2018).

Corroborando Tenopir, Birch e Allard (2012), Harris (2016) destaca que o SSGDP é um “[...] conjunto de serviços que auxilia o pesquisador em todas as fases do ciclo de vida dos dados da pesquisa”. A autora acrescenta que os SSGDPs são considerados distintivos em excelência e qualidade dentro de uma biblioteca (HARRIS, 2016, p. 609, tradução nossa).

Esses serviços são importantes, pois, fornecem suporte para os pesquisadores melhorarem suas práticas e gerenciem melhor seus dados (KERBY, 2016).

2.12.1. Desenvolvimento de serviços: caminhos

Ao se projetar, avaliar ou revitalizar um SSGDP, é necessário ter uma ideia geral de qual tipo de serviço é preciso, além de ter ampla compreensão dos vários contextos interligados, como a missão da biblioteca, os princípios da comunidade de dados e as tecnologias. Dessa forma, é possível atender às necessidades dos usuários. Além disso, o estabelecimento de SSGDPs exige nova visão tanto das políticas, procedimentos e valores organizacionais estabelecidos quanto das funções centrais da biblioteca e dos seus princípios de acesso imparcial, privacidade, preservação e usabilidade dos dados. As metas dos serviços devem sempre ser orientadas pelas necessidades dos usuários (GERACI; HUMPHREY; JACOBS, 2012).

Além de orientadas ao usuário, as atividades desenvolvidas nos SSGDPs são norteadas pelas fases do ciclo de pesquisa. Jones, Pryor e Whyte (2013) apontam três fases distintas:

1. pré-concessão: assistência com o plano de gestão de dados;
2. ao longo do projeto: aconselhamento, orientação e assessoria para documentação de dados, desenvolvimento de formatos e padrões para permitir a reutilização; armazenamento, gestão e análise de dados; assessoria e/ou fornecimento de instalações de armazenamento que atendam às necessidades de uma ampla variedade de tipos de dados, plataformas e necessidades de acesso;
3. pós-projeto: aconselhamento sobre a seleção de dados de valor a longo prazo; apoio para tornar os dados da pesquisa visíveis e/ou disponíveis; ajuda para arquivar dados - no final ou em qualquer outra fase do projeto.

Os autores apontam que os papéis e responsabilidades na definição e fornecimento de SSGDPs não podem ser delegados a uma única unidade, como só as bibliotecas, mas a três grupos (JONES; PRYOR; WHYTE, 2013):

- a) gestores da universidade, visando apoiar e garantir que os serviços propostos sejam desejáveis, realizáveis e sustentáveis, tratando os SSGDPs como investimento sério em infraestrutura dentro de um planejamento institucional de longo prazo;

- b) equipes de suporte e serviços administrativos - incluindo bibliotecas, equipes de tecnologia da informação (TI), que trabalhem em esforço coeso, garantindo a identificação de requisitos, padrões, soluções, práticas de gestão. Atuem, também, no desenvolvimento e implementação de infraestruturas tecnológicas e humanas; em treinamentos de equipes, pesquisadores e gestores; e na divulgação, explicação e promoção de serviços;
- c) pesquisadores, são criadores e usuários dos dados de pesquisa. Garantem envolvimento e apoio efetivo desse grupo na definição dos requisitos dos SSGDPs, uma vez que a prestação de serviços de apoio à gestão de dados deve ser baseada no entendimento das pesquisas, seus padrões e cronogramas, motivações e prioridades.

Dessa maneira, as instituições devem coordenar os atores e processos para fornecendo as infraestruturas tecnológicas e humanas necessárias (JONES; PRYOR; WHYTE, 2013). Entretanto, há escassez de orientação prática para o desenvolvimento de serviços de dados (COATES, 2014).

Segundo Akers *et al.* (2014, p. 172, tradução nossa), “[...] os caminhos adotados em cada instituição são únicos e definidos por pressões e necessidades específicas do local”, não havendo receita pronta para o desenvolvimento dos serviços, mas atividades comuns no processo como

a identificação dos recursos institucionais disponíveis (por exemplo, varredura ambiental), a escolha de abordagem, a identificação das necessidades dos pesquisadores (por exemplo, análise de requisitos), a identificação de parceiros e colaboradores, a condução de pilotos e a construção de uma comunidade. Depois de investir em serviços que atendam às necessidades de alta prioridade, a manutenção contínua envolve a avaliação de serviços, divulgação, promoção e a manutenção de relacionamentos (COATES, 2014, p. 53, tradução nossa).

Dessa forma, esses serviços não são iguais nas instituições. Variam em razão da instituição a que pertencem, do tipo de biblioteca e dos interesses, habilidades e funções dos bibliotecários (SWANSON; RINEHART, 2016).

Liu e Ding (2016) afirmam que a maior compreensão das necessidades dos pesquisadores melhorará a abordagem e a natureza dos serviços de dados oferecidos aos pesquisadores.

Jones, Pryor e Whyte (2013) indicam que o primeiro passo para o desenvolvimento de SSGDPs é o **reconhecimento das diferenças na cultura organizacional**. É importante definir o contexto institucional para entender quais são os desafios específicos e quais serviços podem resolvê-los. Coates (2014) complementa que o desenvolvimento de

SSGDPs exige a compreensão das várias culturas dos pesquisadores, crucial para o fornecimento de serviços de dados relevantes e utilizáveis.

Outro passo importante no processo de desenvolvimento de SSGDPs é a **identificação dos requisitos**. Ela pode ser realizada, por exemplo, por meio de ferramentas com o *Data Asset Framework* (DAF), uma metodologia baseada em entrevistas para investigar os dados gerados pelos grupos de pesquisa e como eles são gerenciados; e **analisar as lacunas**, imprescindíveis de serem identificadas para a provisão de serviços e excelência que se quer encontrar. A análise de lacunas pode ser realizada por meio da ferramenta *Collaborative Assessment of Research Data Infrastructures and Objectives* (CARDIO) que ajuda a estabelecer consenso sobre as capacidades de gestão de dados de pesquisa e identificar lacunas na provisão atual de serviços (JONES; PRYOR; WHYTE, 2013).

Para além das capacidades ou aspectos que devem ser incluídos no desenvolvimento de SSGDPs, Jones, Pryor e Whyte (2013) sugerem que se criem serviços piloto antes da implementação, garantindo que os SSGDPs sejam adequados aos usuários e aos objetivos da instituição.

2.12.2. Elementos ou componentes dos SSGDPs

Para apoiar a gestão e compartilhamento efetivo de dados, a instituição precisa de estratégia coerente de acordo com o conjunto de serviços. Jones, Pryor e Whyte (2013) propõem um diagrama (FIG. 12) para mostrar os componentes que devem ser abordados no fornecimento de serviços, sendo:

- a) políticas e estratégias para gestão de dados de pesquisa;
- b) planos de negócio e sustentabilidade;
- c) orientação, treinamento e suporte;
- d) planejamento da gestão de dados;
- e) gestão de dados ativos;
- f) seleção e entrega de dados;
- g) repositórios de dados;
- h) catálogo de dados.

FIGURA 12 - Componentes do serviço de suporte à gestão de dados de pesquisa

Fonte: (JONES; PRYOR; WHYTE, 2013, p. 5, tradução nossa).

2.12.2.1. Políticas e estratégias para gestão de dados de pesquisa

O primeiro componente é o desenvolvimento de políticas e estratégias para a gestão de dados de pesquisa. Jones, Pryor e Whyte (2013, p. 6, tradução nossa) afirmam que “ter uma estratégia abrangente é essencial para garantir que os serviços de RDM [SSGDPs] se desenvolvam de forma coerente”. Esses autores apontam três etapas principais para a definição da estratégia: 1) entender sua posição atual - contexto, missão da instituição, fatores internos e externos que influenciam a gestão e compartilhamento de dados como códigos de conduta para pesquisa, políticas de financiadores, legislação nacional e internacional e acordos de colaboração; 2) definir meta futura, identificando as expectativas e analisando as lacunas para identificar as atividades necessárias e; 3) mapear um programa de atividades para fazer essa transição e entregar infraestruturas e serviços.

Além da estratégia, Jones, Pryor e Whyte (2013) apontam que se deve elaborar uma política baseada em fatores externos e, no contexto local, visando estabelecer o núcleo dos

serviços. As políticas fornecem clareza sobre o que é esperado pela instituição e especifica os responsáveis por cada atividade.

Chiware e Mathe (2016) reforçam que um dos elementos mais importantes para o fornecimento de SSGDPs é o desenvolvimento de políticas:

uma estrutura de políticas é importante em um ambiente como esse, pois as questões de gestão de dados de pesquisadores são muito sensíveis e contestadas. No processo de desenvolvimento de políticas, é importante indicar claramente os papéis e responsabilidades dos diferentes atores institucionais e, mais especialmente, da biblioteca, que muitas vezes é vista como mero provedor de recursos de informação (CHIWARE; MATHE, 2016, p. 3-4, tradução nossa).

2.12.2.2. Planos de negócio e sustentabilidade

De acordo com Jones, Pryor e Whyte (2013), o desenvolvimento de SSGDPs representa um empreendimento que exige mudanças organizacionais e comportamentais significativas. Para que a gestão de dados de pesquisa seja incorporada como serviço contínuo e permanente na instituição, é necessário planejamento de sustentabilidade por meio de um plano ou caso de negócio (HIOM *et al.*, 2015).

Os planos de negócios representam a oportunidade de obter aprovação e concordância da administração, garantindo recursos para manutenção de serviços, devendo “[...] estabelecer objetivos, custos previstos e despesas planejadas, implantação e aprimoramento de recursos, o programa de mudanças futuras e benefícios esperados” (JONES; PRYOR; WHYTE, 2013, p. 7, tradução nossa). Esses planos devem ser realizados em fases ou etapas (HIOM *et al.*, 2015; JONES; PRYOR; WHYTE, 2013).

Os planos de negócio devem descrever os retornos dos investimentos em termos de eficiência, custos, impacto da pesquisa, oportunidades de novas pesquisas e aumento de financiamento (JONES; PRYOR; WHYTE, 2013). Dessa maneira, são necessários argumentos persuasivos, considerando a relação custos x riscos, demonstrando benefícios mais difíceis de quantificar, por exemplo, o compartilhamento de dados que pode levar a um aumento na colaboração de pesquisa (HIOM *et al.*, 2015). Os planos de negócio devem abordar questões de sustentabilidade e os custos associados em longo prazo (JONES; PRYOR; WHYTE, 2013).

2.12.2.3. Orientação, treinamento e suporte

São apontados vários níveis de orientação, treinamento e suporte oferecidos a diferentes públicos como pesquisadores, alunos e bibliotecários em todas as áreas do conhecimento. Esses serviços incluem suporte mais genérico como páginas da *web* voltadas para a gestão

de dados de pesquisa, e, também, sessões de consultoria individuais. (JONES; PRYOR; WHYTE, 2013).

As páginas da *web* oferecem orientação e *helpdesk* englobando todo o ciclo de vida da pesquisa, desde o financiamento, criação e gestão de dados até sua preservação e reutilização, direcionando os pesquisadores para o suporte local, quando necessário. Além de orientação e *helpdesk*, as instituições podem oferecer treinamentos para os diversos públicos, incluindo formação para a equipe de suporte local e serviços de consultoria, que fornece suporte mais prático e personalizado aos pesquisadores (JONES; PRYOR; WHYTE, 2013).

Todos os serviços devem englobar todo o ciclo de vida dos dados de pesquisa (JONES; PRYOR; WHYTE, 2013).

2.12.2.4. Planejamento da gestão de dados

O planejamento da gestão de dados é esforço colaborativo que integra pesquisadores, bibliotecários ou outros que sejam responsáveis pela preservação e gestão de dados em longo prazo, questões políticas, financeiras *etc.* (JONES; PRYOR; WHYTE, 2013).

Planejar a gestão de dados traz benefícios aos pesquisadores, economizando tempo e esforço, permitindo tomadas de decisões com base em informações, antecipando e evitando problemas (JONES; PRYOR; WHYTE, 2013).

O planejamento de gestão de dados é estabelecido na forma de Plano de Gestão de Dados (PGD), um documento formal que traz “[...] um resumo básico dos dados a serem coletados e explica como eles serão gerenciados, compartilhados e preservados [...]” (JONES; PRYOR; WHYTE, 2013, p. 11, tradução nossa).

O “[...] plano de gestão de dados evolue a partir de um entendimento de como os dados devem ser coletados, normalizados, processados, analisados, preservados, usados e reutilizados durante sua vida útil” (NNLM, 2018, *online*, tradução nossa).

Além de representar conformidade com a política dos financiadores e instituições, o planejamento da gestão de dados, na forma de Plano de Gestão de Dados (PGD) cria oportunidade para os pesquisadores demonstrarem conscientização das melhores práticas de gestão, assegurando que os dados sejam gerenciados de acordo com as políticas estabelecidas (JONES; PRYOR; WHYTE, 2013).

O PGD deve incluir informações de forma abrangente sobre os dados como “[...] os tipos de dados, os padrões de metadados usados, as políticas de acesso e compartilhamento e os

planos de arquivamento e preservação de dados para que sejam acessíveis ao longo do tempo”, garantindo que os dados estarão documentados e disponíveis para reuso (NNLM, 2018, *online*, tradução nossa).

As universidades podem apoiar a criação de PGDs desenvolvendo modelos, orientações e exemplos de PGDs; fornecendo e personalizando ferramentas para elaboração de PGDs; e oferecendo treinamentos, orientações e serviços de consultoria tanto para a concepção do PGD quanto para as ferramentas utilizadas (JONES; PRYOR; WHYTE, 2013).

2.12.2.5. Gestão de dados ativos

A gestão de dados de pesquisa é uma tarefa abrangente em todo o ciclo de vida dos dados. Precisa incluir a gestão dos dados durante a realização da pesquisa, na chamada gestão de dados ativos. Os dados nessa etapa “[...] não são estáticos. Continuam sendo analisados e trabalhados” (SESARTIC; TÖWE, 2016, p. 287, tradução nossa).

Durante a fase ativa da pesquisa, a instituição deverá se preocupar em prover um armazenamento suficiente e *backup* seguro de dados para seus pesquisadores. As soluções de propostas devem oferecer flexibilidade e funcionalidades exigidas pelos pesquisadores para armazenar, acessar e compartilhar dados durante as colaborações da investigação (JONES; PRYOR; WHYTE, 2013). As soluções não devem ser entendidas como ferramentas de publicação ou preservação de dados, mas como fontes para os processos de pesquisa (SESARTIC; TÖWE, 2016).

Várias soluções podem ser implementadas. Mas, uma coleta de requisitos, identificando as práticas de gestão de dados realizadas pela instituição deve antecipar o processo. Isso inclui o local onde os dados são armazenados e a avaliação da infraestrutura e da suficiência dos sistemas (JONES; PRYOR; WHYTE, 2013).

Caso seja necessário, deve-se também oferecer armazenamento adicional para os dados de pesquisa, seja por meio de utilização de instalações de computação de alto desempenho, ampliação das capacidades existentes como pelo oferecimento de opções seguras de armazenamento em nuvem (JONES; PRYOR; WHYTE, 2013).

2.12.2.6. Seleção e entrega de dados

Como não é possível manter todos os dados digitais para sempre, é necessário selecionar os dados para curadoria (WHYTE; WILSON, 2010).

A seleção é um processo “[...] essencial para priorizar dados para curadoria em longo prazo” (JONES; PRYOR; WHYTE, 2013, p. 16, tradução nossa). Apesar do custo e tempo gasto,

selecionar dados para retenção compensa em termos de custo de preservação, armazenamento e disponibilização *online*, comparado ao custo de se manter todos os dados produzidos pela instituição. Dessa forma, deve ser estabelecida uma política para avaliação e seleção dos dados a preservar, garantindo uma tomada de decisão consistente, transparente e responsável. A política deve atender aos requisitos legais relacionados à privacidade e aos direitos de propriedade intelectual, sendo necessário cumprir a legislação relevante para a jurisdição, como políticas nacionais de dados, códigos de conduta adotados pela instituição ou financiador e políticas de governança da informação relacionadas à disciplina (WHYTE; WILSON, 2010).

A política estabelecerá critérios para orientar as decisões sobre o que deve ser mantido pelas instituições. Whyte e Wilson (2010) propõem sete critérios gerais para as decisões de seleção: 1) relevância para a missão (observação das prioridades declaradas na missão da instituição ou na política do órgão de financiamento, incluindo qualquer requisito legal para reter os dados); 2) valor histórico, científico e cultural dos dados; 3) exclusividade (única ou mais completa fonte de informação que pode ser derivada); 4) potencial de redistribuição (confiabilidade, integridade e usabilidade dos arquivos de dados e dos formatos que atendem aos critérios técnicos designados); 5) não replicabilidade dos dados ou recursos ou inviabilidade financeira de replicabilidade; 6) relação custo x benefício; 7) documentação completa (as informações necessárias para futuras descobertas, acessos e reutilizações incluindo metadados de proveniência e de contexto de sua criação e uso).

É necessário incentivar os pesquisadores a selecionarem e depositarem dados por meio de orientação de alto nível, com diretrizes claras sobre: quais dados se enquadram no âmbito do serviço institucional; quais os tipos de dados serão aceitos; e quais as áreas prioritárias (JONES; PRYOR; WHYTE, 2013).

É fundamental *advocacy* defendendo os benefícios do depósito de dados; orientação sobre quais dados são úteis em termos científicos ou acadêmicos; quais informações contextuais, *software* ou *hardware* são necessários; desenvolvimento ou fornecimento de ferramentas de depósito para facilitar o processo de entrega de dados; e estabelecimento de contratos de depósitos visando fixar os termos, condições e responsabilidades tanto dos depositantes quanto dos repositórios (JONES; PRYOR; WHYTE, 2013).

2.12.2.7. Repositório de dados

Após selecionar quais dados manter, deve-se definir onde eles serão preservados e compartilhados. A solução para o depósito de dados deverá levar em conta três opções básicas: 1) desenvolvimento e manutenção de repositório de dados institucional; 2) uso de

repositório externo de dados de pesquisa; e 3) indicação de repositórios ou serviços adequados (JONES; PRYOR; WHYTE, 2013).

A escolha da solução deverá levar em conta o custo para manutenção de um repositório e provisão de pessoal especializado, assim como os custos de preservação em longo prazo. Além disso, deve-se considerar a disponibilidade de serviços existentes, como repositórios específicos de assunto e centros de dados fornecidos por financiadores de pesquisa (JONES; PRYOR; WHYTE, 2013).

As infraestruturas tecnológicas, incluindo os repositórios de dados já foram tratadas em seção anterior.

2.12.2.8. Catálogo de dados

O catálogo de dados permite aos usuários encontrarem, compartilharem e usarem dados (GUPTILL, 1999). De acordo com o autor, uma “[...] coleção de registros de metadados combinados com ferramentas de gestão e pesquisa de dados forma um catálogo de dados” (GUPTILL, 1999, p. 677, tradução nossa). As instituições devem ter registro dos seus dados de pesquisa e disponibilizem os metadados *online* para dar suporte à descoberta e reutilização de dados (JONES; PRYOR; WHYTE, 2013).

Devem-se estabelecer os metadados necessários para registrar os conjuntos de dados de pesquisa, e, ainda, constituir um sistema para capturar e exibir um registro de dados. Sempre que possível, deve-se integrar sistemas, trazendo benefícios à coleta de dados e à criação de metadados em fluxos de trabalho existentes. Finalmente, devem-se expor os metadados para inclusão em catálogos nacionais ou outros serviços (JONES; PRYOR; WHYTE, 2013).

Além dos componentes apontados anteriormente por Jones, Pryor e Whyte (2013), Qin, Crowston e Kirkland (2017) afirmam que, para gerenciar dados de pesquisa, é necessário construir capacidades técnicas, de comunicação, de pessoal e políticas nos níveis institucional e de projetos. Essas capacidades devem avançar gradualmente até que se alcance os níveis de maturidade da capacidade de gestão de dados de pesquisa, partindo do nível 1 - intuitivo, onde os dados são gerenciados intuitivamente em nível de projeto, contudo sem metas e práticas claras, até o nível 5 - otimizado, onde a gestão de dados concentra-se na melhoria do processo. No nível otimizado, os processos relacionados ao fluxo de trabalho para disseminação de dados, incluindo compartilhamento, descoberta e citação, são avaliados regularmente. Melhorias necessárias são implementadas.

Chiware e Becker (2018) assim como Qin, Crowston e Kirkland (2017) afirmam que o desenvolvimento de SSGDPs explora questões relacionadas à qualificação de pessoal e vão além, indicando que os SSGDPs envolvem infraestruturas, mandatos nacionais e colaborações institucionais entre diferentes serviços de apoio.

2.12.3. Tipos e classificações de serviços

Uma variedade de serviços pode ser ofertada pelas bibliotecas. De acordo com Tenopir *et al.* (2017) a gestão de dados de pesquisa assume muitas formas. Os SSGDPs estão presentes desde a ajuda aos pesquisadores na localização de recursos sobre planejamento da gestão de dados ou padrões de metadados, até a criação e manutenção de repositórios de dados. Além disso, os serviços incluem o fornecimento de ferramentas para mineração e visualização de dados; treinamento sobre as atividades de gestão; orientação sobre políticas institucionais; ajuda na criação de planos de gestão de dados e na criação de metadados; e assistência com questões de propriedade intelectual e de privacidade dos dados.

Para Swanson e Rinehart (2016) os SSGDPs incluem: educação; capacidade técnica para vários componentes do ciclo de vida dos dados; planejamento para gestão de dados, assistência para a coleta de dados a partir de fontes textuais; aconselhamento ou assessoria sobre documentação dos dados e uso de metadados; demonstração do impacto da publicação do conjunto de dados (*dataset*); e fornecimento de ferramentas para descoberta, acesso e preservação do conjunto de dados.

Lyon (2012) aponta 10 serviços:

1. levantamento de requisitos, auditoria para identificação das necessidades em relação aos dados de pesquisa;
2. planejamento, oferecendo *advocacy* e orientação a estudantes e pesquisadores no planejamento efetivo de gestão de dados, além de contribuir com o planejamento anual da instituição em questões sobre futuras demandas de infraestrutura de dados, requisitos de armazenamento, problemas de segurança e sensibilidade e acesso;
3. informática, fornecendo conhecimentos técnicos para a descrição de dados estruturados, incluindo o uso de padrões e esquema de metadados, formatos de dados, ontologias apropriadas para disciplinas e domínios específicos;
4. citação, fornecendo orientação e *links* para serviços de terceiros que permitam a atribuição de identificadores persistentes a conjuntos de dados para suportar descobertas, citações e reutilização;

5. treinamento, colaborando com as instituições para propor programas e módulos de treinamento de gestão de dados para toda a equipe de pós-graduação (pesquisadores e estudantes);
6. licenciamento, indicando orientação especializada para auxiliar nas consultas sobre licenciamento de dados, que abrange questões legais e éticas associadas aos conjuntos de dados;
7. avaliação, fornecendo orientação, avaliação e seleção dos dados de pesquisa que devem ser mantidos;
8. armazenamento, colaborando com os serviços de tecnologia da informação (TI) para garantir a clareza e relevância das diretrizes locais de armazenamento de dados e provisão de infraestrutura;
9. acesso, informando sobre o desenvolvimento de políticas de liberação de dados e embargos que apoiam agendas científicas abertas e *Freedom of Information* (FOI)³⁴, respeitando a propriedade intelectual do pesquisador, da instituição e de todos os parceiros industriais colaboradores. O acesso garante vínculos apropriados de dados adicionais e complementares a repositórios institucionais;
10. impacto, promovendo a vinculação com as agências de apoio à pesquisa para a coleta de provas de impacto diretamente relacionadas aos conjuntos de dados de pesquisa e sua reutilização. As bibliotecas podem promover o uso de serviços de acompanhamento de impacto (em exemplo, total.impact.org) e ajudar os pesquisadores a maximizar seu alcance na *web*.

A inclusão dos SSGDPs nas bibliotecas exige uma revisão de sua estrutura organizacional, serviços de informação e reatribuição de recursos (LYON, 2012).

Os serviços mencionados anteriormente são apenas alguns daqueles que as bibliotecas podem oferecer. Além de apontar serviços, alguns autores propõem níveis, categorias ou classificações para os SSGDPs. Alguns desses agrupamentos são apresentados a seguir.

³⁴ *Freedom of Information* (Liberdade de informação) - promove maior abertura e responsabilidade nos órgãos públicos concedendo "[...] direito geral de acesso público a todas as formas de informações "gravadas" detidas pelas autoridades públicas, estabelecendo isenções desse direito geral e impondo uma série de obrigações para as autoridades públicas" (MCGEEVER; WHYTE; MOLLOY, 2015, p. 7, tradução nossa).

2.12.3.1. Classificação dos serviços

Tenopir, Birch e Allard (2012) afirmam que os serviços podem ser classificados como serviços informativos/consultivos (1) ou serviços técnicos (2). Os serviços informativos (1) incluem a consultoria em planos de gestão de dados e padrões de metadados, suporte de referência para encontrar e citar conjuntos de dados, fornecimento de guias *web* e *help* para dados ou conjuntos de dados. Já os serviços técnicos (2) incluem o suporte técnico para repositórios de dados, preparação de conjuntos de dados para um repositório, a desativação ou desmarcação do conjunto de dados de um repositório ou criação de metadados para um conjunto de dados (TENOPIR; BIRCH; ALLARD, 2012). Os serviços informativos/consultivos são os mais comumente presentes nas bibliotecas. O quadro a seguir (Quadro 15) apresenta os tipos de serviços:

QUADRO 15 - Tipos de serviços oferecidos ou planejados pelas bibliotecas

Serviços informativos ou de consultoria	Serviços técnicos
Consultoria na elaboração de planos de gestão de dados	Suporte técnico para sistemas - repositório, sistemas de acesso e descoberta
Consultoria na definição de padrões de dados e metadados	Remoção de dados ou conjunto de dados de um repositório
Divulgação e cooperação com outros provedores de serviços de dados, dentro ou fora do <i>campus</i>	Preparação de dados ou conjuntos de dados para depósito em um repositório
Suporte de referência para encontrar e citar conjuntos de dados e/ou dados	Criação ou transformação de metadados em dados ou conjuntos de dados
Criação de guias <i>web</i> e ajuda para encontrar conjuntos de dados, dados ou repositórios de dados	Identificação de dados ou conjuntos de dados que poderiam ser candidatos a repositórios dentro ou fora do <i>campus</i>
Participação regular em discussões sobre os serviços de dados com outros bibliotecários e outros profissionais	Participação direta em projetos de serviços de dados como membro da equipe, em conjunto com outros pesquisadores
Treinamento de colegas de trabalho em serviços de dados	

Fonte: Elaborado pela autora com base em Tenopir, Birch e Allard (2012).

Geraci, Humphrey e Jacobs (2012) propõem a classificação de serviços em três níveis principais: 1) serviços de coleção (criação e manutenção de coleções); 2) serviços de referência em dados ou curadoria de dados; e 3) serviços de computação (armazenamento, entrega e uso de dados).

Reznik-Zellen, Adamick e McGinty (2012) sugerem três níveis de serviço: 1) educação (bibliotecas educam sua comunidade sobre gestão de dados); 2) consulta (suporte sobre uma variedade de questões relevantes para a gestão de dados de pesquisa); e 3) infraestrutura (fornecimento de infraestrutura para gestão e curadoria de dados).

Hiom *et al.* (2015) classificam os serviços em: 1) assistência à gestão de dados por meio de aconselhamento e orientação; 2) *advocacy* e treinamento direcionados à todos os docentes, instalações de pesquisa e outros grupos que apoiam a pesquisa dentro da universidade; 3) armazenamento e suporte à publicação de dados; e 4) presença na *web*, mídia social e métricas.

Si *et al.* (2015) dividem os serviços em seis aspectos:

1. introdução aos dados de pesquisa - busca explicar a definição de dados de pesquisa, as razões e os valores de gestão, preservação e compartilhamento de dados de pesquisa, dentre outras;
2. diretrizes do plano de gestão de dados - traz informações sobre a escrita de PGDs que atendam aos requisitos do financiador. Esse serviço inclui uma introdução aos PGDs, seus componentes, documentação ou criação de metadados, requisitos de financiamento e modelos ou ferramentas para criação de PGDs;
3. serviços de armazenamento e curadoria de dados - garantem a gestão dos dados ativos, a preservação de longo prazo e o acesso a dados de pesquisa;
4. treinamentos em gestão de dados - *workshops*, tutoriais, palestras e cursos *online*. Os treinamentos ajudam os pesquisadores a lidarem com os desafios encontrados na gestão de dados durante o processo de pesquisa;
5. referência em gestão de dados - disponibilização de bibliotecário de referência em dados (via *e-mail*, *site*, bate-papo) de forma a ajudar o pesquisador no processo de gerenciar e preservar seus dados de pesquisa;
6. recomendação de recursos - repositórios de dados de pesquisa e ferramentas, inclusive para a criação de plano de gestão de dados.

Yoon e Schultz (2017) examinaram *websites* de bibliotecas acadêmicas americanas e classificaram os seus serviços em quatro áreas: 1) serviço; 2) informação; 3) educação; 4) rede (*links* para recursos externos).

Conclui-se que a classificação de serviços varia bastante, dependendo do entendimento dos autores que a propõem.

Em busca de seu papel na gestão de dados de pesquisa, incluindo as atividades de curadoria e preservação, as bibliotecas foram impulsionadas a reavaliar como poderiam apoiar todo o processo de pesquisa. Isso porque as instituições foram obrigadas a gerenciar

seus dados de pesquisa, principalmente, a partir da imposição das agências financiadoras (LATHAM, 2017). Nesse sentido, surgem diversos desafios para o oferecimento de SSGDPs por parte das instituições. Alguns desses desafios serão abordados na seção seguinte.

2.12.4. Desafios para o oferecimento de SSGDPs nas bibliotecas

As bibliotecas são instituições que sempre desempenharam papel importante na gestão de materiais em formato textual (HENTY, 2008). Agora, elas terão que se reinventar. Para Monastersky (2013) as bibliotecas precisarão fazer *reboot* para acompanhar as rápidas transformações do mundo acadêmico do século XXI.

Essas instituições fazem um trabalho ativo na compra de livros, acesso às revistas acadêmicas, fornecimento de espaço de estudos e treinamento em pesquisa para os estudantes, além de ajudar professores no desenvolvimento de cursos e na adoção de novas tecnologias. Contudo, houve certo distanciamento da biblioteca com o cientista, que permanece em seus gabinetes ou laboratórios de pesquisa, acessando informação em formato *online*. Para muitos cientistas, as bibliotecas já não atendem às suas necessidades (MONASTERSKY, 2013).

Esse cenário pode mudar em breve, com as bibliotecas tornando-se mais ativas junto aos pesquisadores, comenta Monastersky (2013, p. 431, tradução nossa). “As bibliotecas estão procurando ajudar com todas as etapas da pesquisa, oferecendo orientação e ferramentas para colecionar, explorar, visualizar, rotular e compartilhar dados”.

Corroborando Henty (2008), Lyon (2012) argumenta que as bibliotecas tradicionalmente apoiaram o processo de pesquisa e comunicação acadêmica, principalmente por meio do acesso e preservação de resultados de pesquisa publicados. Agora, o cenário da pesquisa está mudando drasticamente. Os dados ocupam o centro do processo. A autora defende que “[...] agora é a hora de analisar criticamente sua estrutura [da biblioteca], função e portfólio de serviços, para garantir que eles estejam aptos a apoiar a pesquisa intensiva em dados de hoje efetivamente” (LYON, 2012, p. 127, tradução nossa).

Apesar de ampla prática na gestão de materiais em formato textual (HENTY, 2008), o papel das bibliotecas ainda é pouco claro, pois elas não têm a experiência para lidar com armazenamento de dados de pesquisa, principalmente em grande escala (GOLD, 2007b). Mesmo assim, elas podem gerenciar dados em colaboração com outras partes interessadas, pois possuem *expertise* para contribuir com a gestão de dados (CORRALL, 2012).

A gestão de dados representa enormes desafios. Segundo Lewis (2010), apesar do trabalho bem sucedido das bibliotecas universitárias, diante de novas demandas de aprendizagem em termos de competência informacional, as bibliotecas precisam reatar o vínculo com o pesquisador. Este, até os anos 90, era muito próximo às bibliotecas, pois essas instituições eram as detentoras da informação. O acesso às fontes era realizado na biblioteca. Assim, a mediação entre as fontes de informação e o pesquisador era realizado pelos bibliotecários. Hoje, o vínculo diminuiu substancialmente em razão do acesso eletrônico às fontes de informação, gerando um distanciamento físico entre bibliotecas e pesquisadores. O sucesso na tarefa de gestão de dados de pesquisa necessita de engajamento destes dois atores: pesquisadores e bibliotecários.

Gold (2007a) aponta outros três grandes desafios à gestão de dados de pesquisa em bibliotecas: o primeiro deles é a alfabetização do bibliotecário em ciberinstrução (1). O profissional deve entender a ciberinfraestrutura, a *eScience*, os colaboradores, a colaboração na ciência, a ciência computacional e de grade, a curadoria de dados, a *web* semântica, os dados abertos, o arquivamento e preservação digital, a gestão de dados, compreendendo como todos esses tópicos se relacionam.

O segundo grande desafio dos bibliotecários é a tomada de consciência sobre as questões da ciência de dados e da própria gestão de dados. As bibliotecas estão fortemente voltadas para os produtos da ciência tradicional – comunicação e publicação científica com orientação textual. Os profissionais da área apresentam grande domínio das fases pós-pesquisa – relatórios, comunicação e publicação – ao passo que estão pouco familiarizados com a geração de dados no ciclo de pesquisa científica (GOLD, 2007a).

O terceiro desafio estaria na gestão de comunicação do processo, pois a linguagem pode se transformar em obstáculos se não adequada. Um vocabulário familiar aos bibliotecários como arquivamento, curadoria, administração e proveniência assumem significado novo ou especializado no contexto da *eScience* (GOLD, 2007a).

Fearon *et al.* (2013) apontam três grandes desafios para o oferecimento dos SSGDPs nas bibliotecas:

- a) financiamento dos serviços, seja por orçamento regular na instituição, financiamento externo ou por meio de projetos temporários;
- b) formação de parcerias para fornecimento de serviços abrangentes e apoio à pesquisa durante todo o ciclo de vida;

- c) envolvimento do corpo docente que desconhece os serviços que a biblioteca oferece, resultando em baixo valor percebido dos serviços, além de resistência ao compartilhamento de dados.

Além desses, Fearon *et al.* (2013) apresentam outros desafios como: desenvolvimento de infraestrutura tecnológica; falta de pessoal dedicado à gestão de dados de pesquisa; falta de plano de *marketing*, comunicação ou conscientização sobre os serviços; treinamento de pessoal; definição dos papéis da equipe; escopo dos serviços (fundamentado nas necessidades dos pesquisadores, com foco na gestão ativa e na preservação em longo prazo); apoio institucional; educação e formação do corpo docente; ausência de políticas institucionais; e ambiguidade dos requisitos das agências financiadoras, uma vez que os serviços normalmente iniciam a partir de uma imposição externa, conforme apontam Tenopir *et al.* (2015).

Lewis (2010) aponta a infraestrutura; as habilidades e conhecimentos da equipe; e a cultura dos pesquisadores como os principais desafios.

Já Goldman, Kafel e Martin (2015) apresentam desafios como a falta de uma política de gestão de dados na instituição; a dificuldade em aumentar a conscientização sobre os serviços oferecidos pelas bibliotecas; a confusão em torno da definição de 'serviços de dados'; a colaboração entre pesquisadores e departamentos; a conscientização sobre a gestão de dados; as mudanças na equipe e na administração; a falta de tempo e conhecimento da equipe; as estratégias ineficazes de promoção do papel da biblioteca; a não disposição dos pesquisadores para o compartilhamento de dados; e o tamanho e falta de especialização da equipe.

Cox, Verbaan e Sen (2012) apontam desafios como: redução de outras prioridades da equipe em detrimento dos papéis relacionados à gestão de dados; falta de experiência em pesquisa por parte do bibliotecários; a compreensão da diversidade de dados de pesquisa e das práticas de pesquisa dentro das culturas e subculturas disciplinares; conhecimento específico de domínio; problemas em manter um papel ativo junto ao pesquisador; problemas relacionados à escala, apoio e aconselhamento genérico em devida às diferenças disciplinares, práticas e dados; complexidade e escala dos problemas de gestão de dados nas instituições; recursos, infraestrutura, políticas e estruturas de governança ainda em mudança contínua).

Raboin, Reznik-Zellen e Salo (2012) apontam a captação de apoio administrativo; a continuidade do atendimento às necessidades dos pesquisadores uma vez que elas evoluem; a captação de número maior de funcionários da biblioteca para as atividades de

gestão de dados; e a gestão e integração de serviços com estruturas de pessoal novas ou existentes.

Soehner, Steeves e Ward (2010) listam como desafios o apoio institucional; o estabelecimento de colaboração (parcerias) dentro da instituição; o envolvimento do corpo docente; a infraestrutura tecnológica necessária; o conhecimento limitado da equipe; a falta de políticas para a gestão de dados; e a falta de financiamento e recursos.

Creamer *et al.* (2012) indicam como desafios o financiamento para os serviços (pessoal e equipamentos); as lacunas de habilidades e conhecimentos da equipe; a falta de apoio institucional; e as lutas territoriais entre os vários departamentos da instituição.

Brown (2010) indica como desafios a sustentabilidade dos serviços (financiamento); a qualificação e treinamento dos bibliotecários; e a comunicação e compreensão correta da importância da gestão de dados.

Em síntese, há uma concordância entre os autores sobre os principais desafios para o oferecimento de SSGDPs pelas bibliotecas. Esses desafios são, muitas vezes, enormes. Cabe às instituições, mesmo diante desses desafios e barreiras, continuarem a fornecer um conjunto de serviços que atendam às necessidades da sua comunidade.

O próximo capítulo apresenta os procedimentos metodológicos utilizados para se atingir os objetivos propostos nesta tese.

3. PROCEDIMENTOS METODOLÓGICOS

Nesta etapa serão apresentados os procedimentos metodológicos propostos para esta investigação, integrando tanto a teoria quanto o método de pesquisa, de forma que o pesquisador debata suas ideias, opções e práticas.

3.1. CARACTERIZAÇÃO DA PESQUISA

Conforme classificação apresentada por Gil (1994), quanto à sua natureza, a pesquisa pode ser considerada **aplicada**, pois apresenta um “[...] propósito prático e específico para gerar conhecimento ou avanço do conhecimento e tem uma esfera de atuação mais abrangente, seja científica, industrial, econômica ou acadêmica” (PERDIGÃO, 2012, p. 110).

Quanto à forma de abordagem do problema, as pesquisas podem ainda ser classificadas em pesquisa quantitativa e pesquisa qualitativa (GIL, 1994). Para atingir os objetivos propostos neste estudo adotamos a **abordagem qualitativa**.

Quanto aos seus objetivos, as pesquisas podem ser classificadas em exploratórias, descritivas e explicativas. A pesquisa exploratória proporciona maior familiaridade com o tema, tornando-o mais explícito ou constituindo hipóteses. Esses estudos objetivam o aprimoramento de ideias e a descoberta de intuições. A pesquisa descritiva objetiva a descrição das características de uma população ou fenômeno ou o estudo da relação entre as variáveis. As pesquisas explicativas buscam identificar os fatores que determinam ou que contribuem para a ocorrência dos fenômenos, aprofundando o conhecimento da realidade, explicando a razão, o porquê das coisas (GIL, 1994).

Este estudo, quanto aos seus objetivos, se enquadra em duas classificações: **exploratória**, pois o problema de pesquisa ainda é pouco estudado em literatura nacional; e também **descritiva** por levantar e caracterizar um determinado assunto por meio da literatura.

Além disso, de acordo com Creswell (2010) há três componentes envolvidos no projeto de pesquisa: a **concepção filosófica**, a **estratégia de investigação** que deve ser relacionada à concepção escolhida e os **métodos ou procedimentos de pesquisa** específicos que transformam a abordagem em prática.

Creswell (2010) admite que, embora as concepções filosóficas permaneçam ocultas, elas influenciam a prática de pesquisa e, por isso, devem ser identificadas. O autor considera que as concepções são um conjunto de crenças que guiam a ação. Nesse sentido Creswell (2010) apresenta quatro concepções filosóficas: pós-positivista; construtivista social; reivindicatória e participatória; e pragmática.

Este trabalho opta por usar a **concepção filosófica pragmática**, que se concentra nos problemas de pesquisa e não nos métodos. Surge mais das ações, das situações e das consequências do que nas condições antecedentes - como na concepção pós-positivista. Entende-se que a concepção pragmática utiliza-se de todas as abordagens disponíveis para solucionar os problemas. Nela, os pesquisadores têm mais liberdade de escolha dos métodos, técnicas e procedimentos que melhor se ajustem às necessidades e propósitos do estudo (CRESWELL, 2010).

Uma vez escolhida a abordagem para tratar o problema – abordagem qualitativa - parte-se para a escolha dos métodos ou procedimentos específicos de pesquisa. Michel (2009, p. 50) argumenta que o método é “[...] o plano geral, norteador do processo, o caminho, o modo escolhido para se chegar a uma resposta, a uma solução”, envolvendo, segundo Creswell (2010, p. 40), a “[...] forma de coleta, análise e interpretação dos dados que os pesquisadores propõem para seus estudos”. As próximas seções apresentam os caminhos escolhidos para esta pesquisa.

3.2. ETAPAS DA PESQUISA

Esta pesquisa constou de três etapas distintas:

- 1) identificação de serviços e ações e dificuldades por meio da literatura;
- 2) identificação e contextualização de serviços e identificação das ações e dificuldades por meio de entrevistas;
- 3) análise dos dados:
 - a) sistematização e proposição de uma taxonomia de serviços (literatura e entrevistas);
 - b) compilação de serviços (literatura e entrevistas);
 - c) contextualização dos serviços nas instituições (entrevistas);
 - d) síntese das dificuldades e desafios em relação à implantação dos serviços de suporte à gestão de dados de pesquisa (literatura e entrevistas);
 - e) síntese das ações gerais para a implementação e manutenção dos serviços (literatura e entrevistas).

A seguir, serão descritas cada uma destas etapas, bem como seus objetivos.

3.2.1. Primeira etapa: identificação de serviços e ações

A primeira etapa busca atender ao objetivo de:

- 1) Identificar, por meio de revisão de literatura:

- a) os serviços voltados à gestão de dados de pesquisa promovidos por bibliotecas acadêmicas em âmbito mundial;
- b) as ações necessárias para viabilização e manutenção desses serviços;
- c) as dificuldades e desafios em relação à implantação dos serviços de suporte à gestão de dados nas instituições.

Para atender ao objetivo adotou-se como procedimento técnico a pesquisa bibliográfica em materiais já publicados. Foi realizado levantamento de artigos publicados em periódicos, cuja temática é o suporte à gestão de dados de pesquisa realizado nas bibliotecas universitárias.

A seleção dos documentos que compõem esta etapa foi realizada por meio de pesquisa nas seguintes bases de dados da área de Ciência da Informação: *Library and Information Science Abstracts* (LISA), *Library, Information Science and Technology Abstracts* (LISTA), ambas de cobertura internacional e Base de Dados Referencial de Artigos de Periódicos em Ciência da Informação (BRAPCI) de cobertura nacional, além do Google Acadêmico e referências citadas nos artigos.

A pesquisa nas bases de dados utilizou os termos, presentes nos campos título, resumo e palavras-chave dos artigos:

("data management" AND librar*) OR ("research data servic*" OR "data management servic*" OR "research data support*" OR "data management support*" OR "research data management")

A seleção dos artigos obedeceu aos seguintes critérios:

- a) tipo de publicação: periódico científico;
- b) idioma: português, inglês ou espanhol;
- c) disponibilidade: texto completo;
- d) pertinência: suporte à gestão de dados de pesquisa realizado nas bibliotecas universitárias (cauda longa da ciência).

Após a seleção inicial pelo tipo de publicação, idioma e disponibilidade de texto completo, 203 artigos passaram por leitura do título e resumo visando avaliar sua pertinência. Desses 38 foram descartados.

Os artigos foram importados para o *software* de pesquisa qualitativa NVivo, da QSR International. Todo o conteúdo foi codificado em nós que representam as categorias,

subcategorias e subdivisões de serviços. O processo de criação de nós e codificação dos documentos será tratado na seção **3.2.3 Análise de dados**.

Para esta etapa, o *corpus* utilizado foi de 165 artigos (APÊNDICE C). A quantidade de artigos foi considerada suficiente para cumprir o objetivo da etapa 1, baseado no critério da saturação teórica, conforme tratada por Fontanella *et al.* (2011). A leitura e codificação dos artigos foi interrompida quando se constatou que não havia novos elementos ou novos serviços para a inclusão.

3.2.2. Segunda etapa: identificação e contextualização de serviços, ações e dificuldades

A segunda etapa da pesquisa teve por objetivo:

1. identificar e contextualizar, por meio da visão dos atores institucionais:
 - a) os serviços voltados à gestão de dados de pesquisa oferecidos pela instituição.
2. identificar, por meio da visão dos atores institucionais:
 - a) as ações para viabilização e manutenção desses serviços;
 - b) as dificuldades e desafios em relação à implantação dos serviços de suporte à gestão de dados de pesquisa nas instituições.

A segunda etapa da pesquisa foi realizada em instituições levantadas através de uma amostragem não probabilística intencional. Levando-se em consideração que a amostragem em pesquisa qualitativa não é representativa do universo, a seleção pode ser intencional, de forma a explorar os espectros de opiniões e as diferentes representações sobre o assunto em questão (OLIVEIRA, 2012). Complementando, Gil (1994, p. 104) aponta que a amostragem por tipicidade ou intencional consiste em “[...] selecionar um subgrupo da população que, com base nas informações disponíveis, possa ser considerado representativo de toda a população”. Esta etapa utilizou uma amostra não probabilística intencional.

Minayo (2017, p. 4) acrescenta que a amostra em uma pesquisa qualitativa deve estar vinculada à dimensão do objeto que “[...] se articula com a escolha do grupo ou dos grupos a serem entrevistados [...]” e que, dentre outras recomendações, a escolha dos participantes deve levar em consideração o grupo social relevante, contemplando as características, experiências e expressões e garantindo a complementaridade das informações, de acordo

com os atributos que o investigador pretende conhecer de forma que o quadro empírico da pesquisa esteja mapeado e compreendido.

A seleção dos participantes foi realizada considerando sua representatividade e variabilidade, garantindo diferentes representações sobre o assunto.

Para esta etapa, foi realizada pesquisa na *web* por iniciativas e eventos voltados para a gestão de dados nas universidades brasileiras, buscando identificar instituições envolvidas com a disponibilização de serviços ou infraestrutura para a gestão de dados.

Foram identificadas algumas iniciativas (Quadro 16), sendo:

QUADRO 16 – Iniciativas de gestão de dados nas instituições de ensino superior brasileiras

Instituição	Descrição
Universidade de São Paulo (USP)	Oferece no <i>site</i> do Sistema Integrado de Bibliotecas (SIBi), guia 'Apoio ao Pesquisador', informações sobre dados de pesquisa (plano de gestão de dados, como organizar dados, como citar dados, repositórios de dados e as iniciativas na USP).
Universidade Estadual Paulista "Júlio de Mesquita Filho" (UNESP)	Organizou, em junho de 2018, o seminário Gerenciamento de Dados de Pesquisa no contexto da <i>Open Science</i> .
Universidade Federal de Santa Catarina (UFSC)	Organizou em setembro de 2017 o I Seminário de Suporte à Pesquisa e Gestão de Dados Científicos: panorama atual e desafios.
Universidade Federal de São Carlos (UFSCar)	Disponibiliza informações sobre a gestão de dados de pesquisa na página de sua Pró-reitoria de Pesquisa.
Universidade Federal de São Paulo (UNIFESP)	Organizou em agosto de 2018, juntamente com a UFABC, o I Seminário Gestão de Dados de Pesquisa Unifesp e UFABC.
Universidade Federal do ABC (UFABC)	Organizou em agosto de 2018, juntamente com a UNIFESP, o I Seminário Gestão de Dados de Pesquisa Unifesp e UFABC.
Universidade Federal do Rio Grande (FURG)	Participa do projeto de pesquisa Rede de Dados de Pesquisa Brasileira (RDP Brasil), desenvolvido pela Universidade Federal do Rio Grande do Sul (UFRGS) em parceria com a Universidade Federal do Rio Grande (FURG), sob coordenação executiva da Rede Nacional de Ensino e Pesquisa (RNP) e do Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT).
Universidade Federal do Rio Grande do Sul (UFRGS)	Participa do projeto de pesquisa Rede de Dados de Pesquisa Brasileira (RDP Brasil), desenvolvido pela Universidade Federal do Rio Grande do Sul (UFRGS), em parceria com a Universidade Federal do Rio Grande (FURG), sob coordenação executiva da Rede Nacional de Ensino e Pesquisa (RNP) e do Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT).

Fonte: Dados de pesquisa (2019).

Para esta segunda etapa da pesquisa selecionou-se como casos as instituições identificadas no Quadro 16. Dessa forma foi enviada, por *e-mail*, carta-convite às bibliotecas ou coordenadores dos sistemas de bibliotecas convidando-os para participarem da pesquisa. Foram pedidas informações a respeito da disponibilização e participação das bibliotecas em discussões para o desenvolvimento de serviços ou infraestruturas.

O Quadro 17 apresenta a situação dos contatos com essas instituições.

QUADRO 17 – Respostas à carta-convite e pedido de informações

Instituição	Situação
Universidade de São Paulo (USP)	Não respondeu ao contato.
Universidade Estadual Paulista "Júlio de Mesquita Filho" (UNESP)	A Coordenadoria Geral de Bibliotecas informou que a UNESP participa do grupo de estudos sobre gestão de dados da FAPESP, juntamente com outras instituições do Estado de São Paulo. Informou ainda que, atualmente, estão trabalhando na estruturação do repositório de dados que se encontra em fase de testes. Além disso, a Coordenadoria está programando videoconferências para que as bibliotecas possam se inteirar sobre o assunto. Não indicou um responsável para relatar a experiência da instituição.
Universidade Federal de Santa Catarina (UFSC)	A coordenadora do Sistema de Bibliotecas relatou que os serviços estavam em desenvolvimento e indicou um responsável para relatar a experiência na instituição.
Universidade Federal de São Carlos (UFSCar)	A biblioteca informou que ainda não disponibiliza serviços para gestão de dados de pesquisa. A discussão sobre Repositório de Dados está sendo feita pelo Departamento de Produção Científica do Sistema Integrado de Bibliotecas da UFSCar. Indicou o responsável para relatar a experiência na instituição, contudo não foi possível realizar a entrevista devido à agenda do responsável.
Universidade Federal de São Paulo (UNIFESP)	A UNIFESP participa do grupo de estudos sobre gestão de dados da FAPESP. Está desenvolvendo um repositório de dados. Apesar da disponibilidade em participar da pesquisa, não foi possível realizar a entrevista por questões de agenda do responsável.
Universidade Federal do ABC (UFABC)	Não respondeu ao contato.
Universidade Federal do Rio Grande (FURG)	Não respondeu ao contato.
Universidade Federal do Rio Grande do Sul (UFRGS)	Foi feito contato com um dos integrantes do Grupo de Trabalho da Rede de Dados de Pesquisa Brasileira (RDP Brasil), que também participa de um grupo que trabalha no projeto da UFRGS para infraestrutura de dados de pesquisa da Universidade.

Fonte: Dados de pesquisa (2019).

A entrevista de forma semiestruturada foi o instrumento utilizado para coleta de dados nesta etapa da pesquisa. Para Gil (1994), esta técnica é bastante flexível, pois permite obter informações com profundidade acerca do que as pessoas sabem, esperam e desejam em relação ao problema abordado.

Foi elaborado um roteiro para a entrevista semiestruturada, refletindo o propósito do estudo e desta etapa de pesquisa, conforme Apêndice D. A entrevista semiestruturada não segue, porém, um roteiro inflexível. O roteiro foi adaptado, pois as duas instituições estão com os serviços e infraestruturas em fase de desenvolvimento. Além disso, foi admitido que o roteiro fosse adaptado de acordo com os rumos do diálogo entre o entrevistador e o participante, de forma que a conversa se tornasse mais natural possível.

Participaram desta etapa somente UFSC e UFRGS, pois as demais instituições identificadas não responderam aos contatos, não indicaram um responsável para relatar a experiência ou

o responsável não teve agenda disponível no período da pesquisa, conforme situação apresentada no Quadro 17. Dentre as instituições participantes da pesquisa, na UFSC, foi entrevistado um bibliotecário. Na UFRGS foi entrevistado um docente. Ambos estavam envolvidos com a discussão e desenvolvimento de serviços e infraestruturas para gestão de dados de pesquisa.

As entrevistas foram realizadas pelo pesquisador, em abril de 2019, utilizando o Skype, *software* que permite realizar chamadas de áudio e de vídeo via *internet*, bem como gravar as chamadas.

Na etapa 2, as entrevistas gravadas, que formam o *corpus* da etapa, foram transcritas com foco na informação verbalizada pelos participantes.

A transcrição das informações verbais foi realizada pelo próprio pesquisador, utilizando a ferramenta *online* oTranscribe (<https://otranscribe.com/>). A transcrição foi feita respeitando as regras da gramática da língua portuguesa. Foram realizados pequenos ajustes na grafia. O objetivo não era a realização de análise linguística, conforme admite Duarte (2004, p. 221). Assim, “[...] frases excessivamente coloquiais, interjeições, repetições, falas incompletas, vícios de linguagem, cacoetes, erros gramaticais [...]” foram corrigidos. Embora numa pesquisa qualitativa com amostragem intencional seja difícil a anonimização dos participantes, os nomes desses foram omitidos nas transcrições, sendo apresentado por meio de código (P1, P2).

A apresentação dos trechos da entrevista neste trabalho foi fundamentada na norma da Associação Brasileira de Normas Técnicas (ABNT) NBR 10520 - Informação e documentação - Citações em documentos, sendo a fala do participante tratada como fala de um autor, assim:

- a) citações diretas de até três linhas são apresentadas entre aspas, no corpo do texto;
- b) citações diretas com mais de três linhas são apresentadas destacadas com recuo de 4 cm da margem esquerda e letra menor que a do texto utilizado e sem as aspas;
- c) o nome do autor/entrevistado (o código) é apresentado, dentro do corpo do texto, em letras maiúsculas e minúsculas e, quando ao final do texto, em letras maiúsculas;
- d) as supressões de texto são apresentadas com reticências entre colchetes [...];
- e) os comentários do pesquisador/pessoa que transcreve são apresentado entre colchetes [].

As entrevistas transcritas foram importadas para o *software NVivo* da *QSR International*, lidas e codificadas em nós (categorias, subcategorias e subdivisões), conforme apresentado na seção a seguir.

3.2.3. Terceira etapa: análise dos dados

A análise de conteúdo proposta por Bardin (1977) foi usada como método para o tratamento e interpretação dos dados coletados nas etapas 1 (literatura) e 2 (entrevistas).

A análise de conteúdo é um dos métodos usados para analisar dados textuais (ELO; KYNGÄS, 2008) e é definida como “[...] método de pesquisa para a interpretação subjetiva do conteúdo dos dados textuais através do processo de classificação sistemática da codificação e identificação de temas ou padrões” (HSIEH; SHANNON, 2005, p. 1278, tradução nossa).

Os 165 artigos (texto completo) e as transcrições das duas entrevistas foram importadas para o *software* Nvivo. Todo o material foi lido integralmente e codificado em nós que representam as categorias, subcategorias e subdivisões de serviços.

As categorias, subcategorias e subdivisões de serviços foram criadas com base no referencial teórico da pesquisa, portanto, utilizando análise de conteúdo dedutiva, abordagem que, geralmente, toma por base os trabalhos anteriores, como teorias, modelos, mapas mentais e revisões de literatura (ELO; KYNGÄS, 2008; GIBBS, 2009); e, também, utilizando abordagem indutiva, em que as categorias surgem a partir dos dados ou temas que emergiram dos artigos analisados, sendo abordagem baseada em dados ou codificação aberta (ELO; KYNGÄS, 2008; GIBBS, 2009).

A codificação dos documentos foi realizada manualmente (em oposição à codificação automática) a partir da seleção de trechos (palavras ou frases) no texto e atribuição de um nó previamente definido (abordagem dedutiva) ou criando-se novo nó (abordagem indutiva).

Os nós foram organizados em uma estrutura hierárquica de forma a refinar as conexões entre os temas. Dessa forma, as categorias, subcategorias e subdivisões (nós e subnós) seguem uma estrutura hierárquica, onde “os códigos que guardam semelhança ou se referem ao mesmo assunto são reunidos sob o mesmo ramo da hierarquia [...]” (GIBBS, 2009, p. 98).

Foi realizada revisão contínua da codificação permitindo o agrupamento e a redução de nós (categorias, subcategorias e subdivisões). O agrupamento de nós foi realizado identificando-se as informações distintas e reunindo as semelhantes ou relacionadas.

Conjuntamente, os dados coletados nas etapas 1 e 2 (literatura e entrevistas) permitiram: a proposição de uma taxonomia de serviços (1); a compilação dos serviços de suporte à gestão de dados de pesquisa oferecidos pelas bibliotecas (2); a exposição dos desafios e dificuldades enfrentadas (4); e a síntese das ações gerais para a viabilização e manutenção dos serviços nas instituições (5).

A etapa 2 (coleta de dados por entrevistas) permitiu a: contextualização dos serviços de suporte à gestão de dados de pesquisa nas instituições entrevistadas (3).

Para a sistematização e proposição de uma taxonomia de serviços, os artigos selecionados (165 artigos – etapa 1) e as duas entrevistas (etapa 2) foram lidas integralmente, codificadas e categorizadas.

A partir da leitura dos artigos e entrevistas foi proposta uma categorização inicial para os serviços e hierarquização de termos. Depois, foi realizado rearranjo das categorias com base na literatura, seguida pela padronização e uniformização de nomes e união das categorias com assuntos que apresentam afinidades temáticas.

Posteriormente, foi realizada descrição do conteúdo das categorias, subcategorias e subdivisões, bem como a descrição estatística, permitindo a identificação dos tipos de serviços mais frequentes.

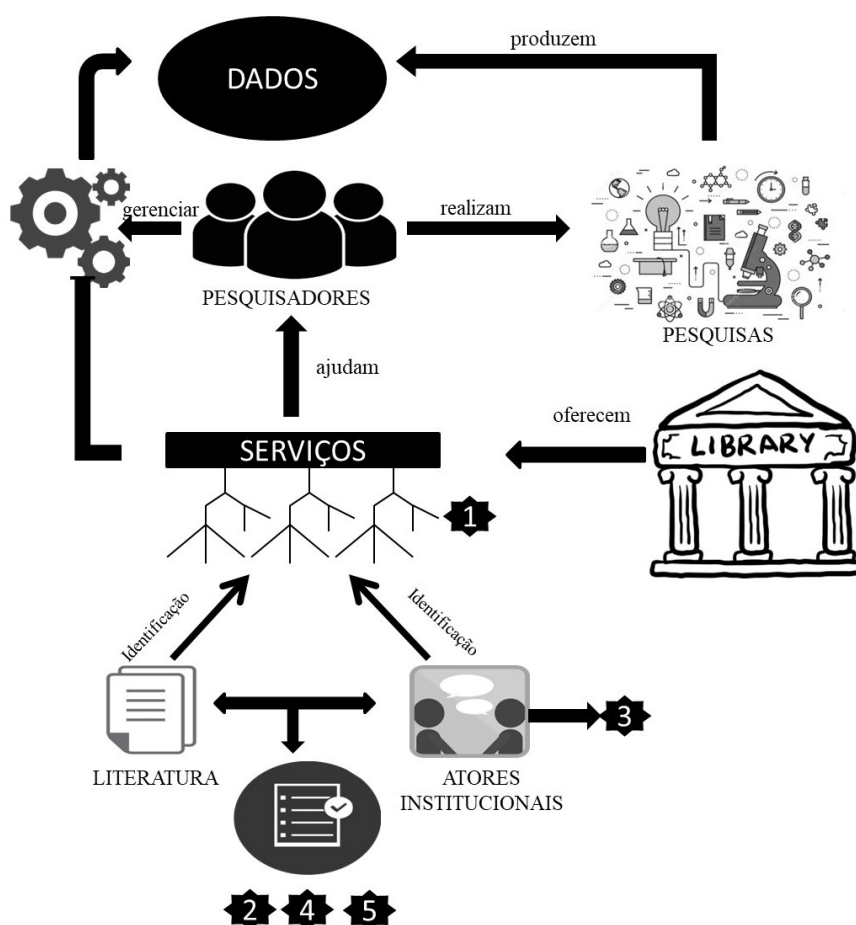
A categorização e a hierarquização dos termos permitiram: sistematizar e propor uma taxonomia de serviços de suporte a gestão de dados de pesquisa, estruturada em três eixos principais – Infraestrutura, Suporte e Educacional (1); e identificar e compilar os SSGDPs (2).

A identificação dos desafios e dificuldades para o fornecimento de SSGDPs (4), bem como a identificação e síntese das ações gerais para a implementação e manutenção dos SSGDPs nas instituições (5) foi elaborada a partir da literatura (165 artigos selecionados) e das duas entrevistas realizadas.

As duas entrevistas realizadas permitiram uma contextualização da atual situação das instituições entrevistadas (3) no que diz respeito ao oferecimento de suporte e infraestrutura para a gestão de dados de pesquisa.

Dessa forma, são elencados os produtos ou resultados deste trabalho, construídos a partir da análise de conteúdo realizada nos artigos selecionados e nas entrevistas. Os números que sucedem os produtos ou resultados (1 a 5) apresentados nos três parágrafos anteriores são representados nesta figura (FIG. 13).

FIGURA 13 – Desenho da pesquisa



Fonte: Elaborado pela autora.

Na parte central da figura, encontram-se os pesquisadores, que realizam pesquisas e produzem grande quantidade de dados. Esses precisam gerenciar seus dados de pesquisa. Para que isso aconteça, as bibliotecas devem oferecer serviços - Serviços de Suporte à Gestão de Dados de Pesquisa (SSGDPs) – que apoiem o pesquisador na gestão de dados de forma adequada. Neste trabalho, a identificação dos SSGDPs foi realizada por meio da literatura e das entrevistas com atores institucionais, possibilitando o primeiro produto da tese: a taxonomia de serviços (1). De modo conjunto, a literatura e as entrevistas proporcionaram mais três produtos: a compilação dos serviços de suporte à gestão de dados de pesquisa (2); a apresentação dos desafios e dificuldades (4); e síntese das ações gerais para viabilização e manutenção dos serviços (5). As entrevistas com atores institucionais permitiram contextualizar os SSGDPs na universidades participantes (3).

O capítulo seguinte apresenta e discute os resultados obtidos, conforme as três etapas de pesquisa detalhadas nos procedimentos metodológicos deste trabalho.

4. APRESENTAÇÃO E DISCUSSÃO DOS RESULTADOS

Esta seção apresenta e discute os resultados obtidos conforme a metodologia proposta. retrata a taxonomia de serviços criada, a análise descritiva das categorias de serviços, a compilação dos serviços de suporte e infraestruturas oferecidos pelas bibliotecas, a contextualização e descrição dos serviços nas instituições brasileiras participantes da pesquisa, as dificuldades enfrentadas para o desenvolvimento dos SSGDPs e a síntese das ações para a viabilização e manutenção dos serviços nas instituições. Apresenta, ainda, as observações, inferências e interpretação desses resultados.

4.1. TAXONOMIA DE SERVIÇOS: APRESENTAÇÃO E ESTRUTURA

A taxonomia de serviços foi criada a partir da leitura de forma integral e categorização de 165 artigos e de duas entrevistas. Foram criados nós - categorias de serviços -, estruturados em três eixos principais – Infraestrutura, Suporte e Educacional, conforme Quadro 18.

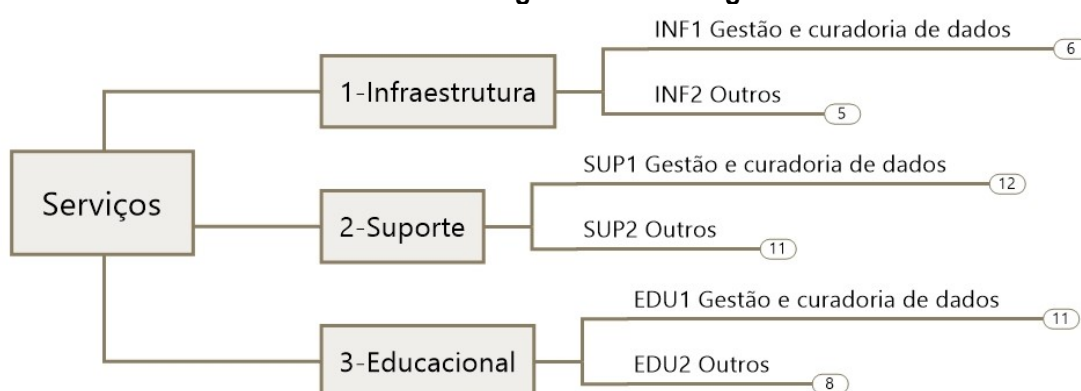
QUADRO 18 – Categorias de serviços

Categorias de serviços	Código
1 – INFRAESTRUTURA	INF
2 – SUPORTE	SUP
3 – EDUCACIONAL	EDU

Fonte: Elaborado pela autora.

As três categorias (Infraestrutura, Suporte e Educacional) são subdivididas nas subcategorias ‘Gestão e curadoria de dados’ e ‘Outros’ conforme a FIG. 14.

FIGURA 14 – Categorias e subcategorias



Fonte: Elaborado pela autora.

Por sua vez, independentemente da categoria, a subcategoria ‘Gestão e curadoria de dados’ (Quadro 19) é formada pelo agrupamento de serviços dispostos em 12 subdivisões. A subcategoria ‘Outros’ é formada por 13 tipos de serviços que não se enquadram nas subdivisões anteriores (Quadro 20).

QUADRO 19 – Subcategoria Gestão e curadoria de dados

Gestão e curadoria de dados	
000	Gestão e curadoria – geral
001	Planejamento
002	Obtenção
003	Organização e formato
004	Documentação de dados e metadados
005	Avaliação e seleção
006	Armazenamento
007	<i>Copyright</i> , propriedade intelectual e licenciamento
008	Questões éticas, legais e de consentimento
009	Publicação e citação
010	Uso e reuso
011	Mineração, análise e visualização

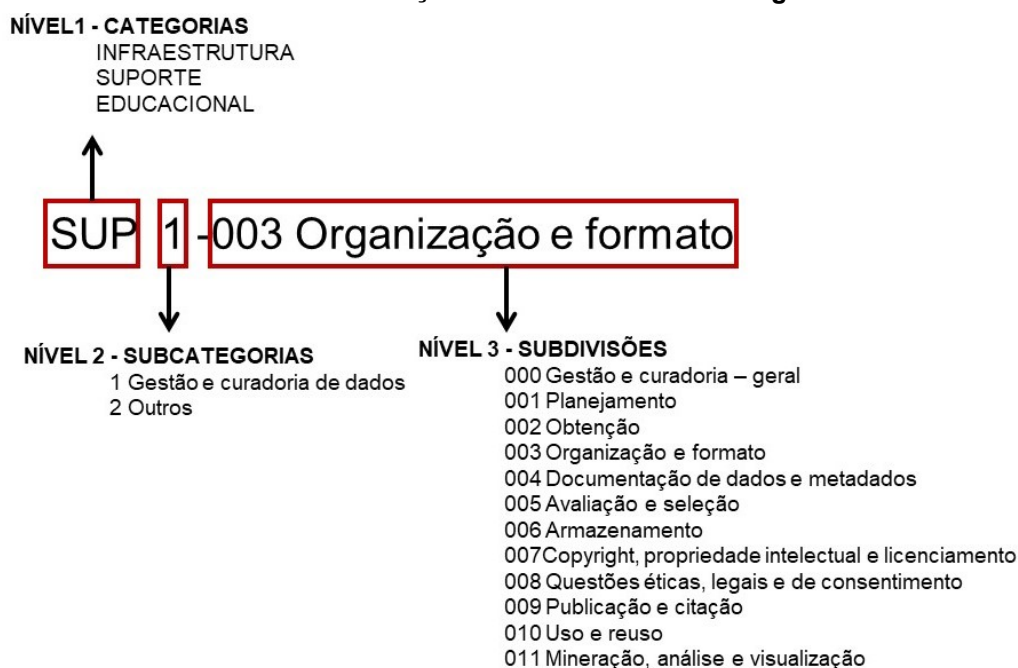
Fonte: Elaborado pela autora.

QUADRO 20 – Subcategoria Outros

Outros	
012	Análise de redes sociais
013	Assistência à produtividade e trabalho colaborativo
014	Concepção e gestão de bases de dados
015	Desenvolvimento de diretrizes, padrões, serviços e infraestruturas
016	Desenvolvimento de disciplina ou <i>curriculum</i>
017	Gestão de pesquisa e comunicação científica
018	Identificação das percepções, práticas e necessidades
019	Participação em projeto ou processo de pesquisa
020	Políticas
021	Rastreamento de impacto
022	Reprodutibilidade da pesquisa
023	Serviços
024	Tecnologias, infraestruturas e ferramentas

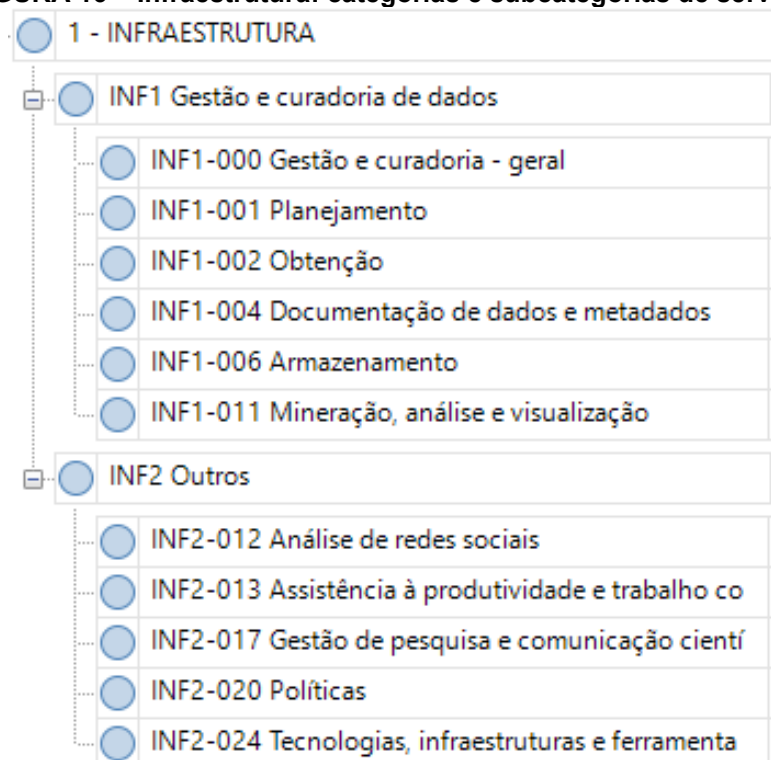
Fonte: Elaborado pela autora.

A nomenclatura das categorias, subcategorias e subdivisões obedece ao esquema a seguir (FIG. 15):

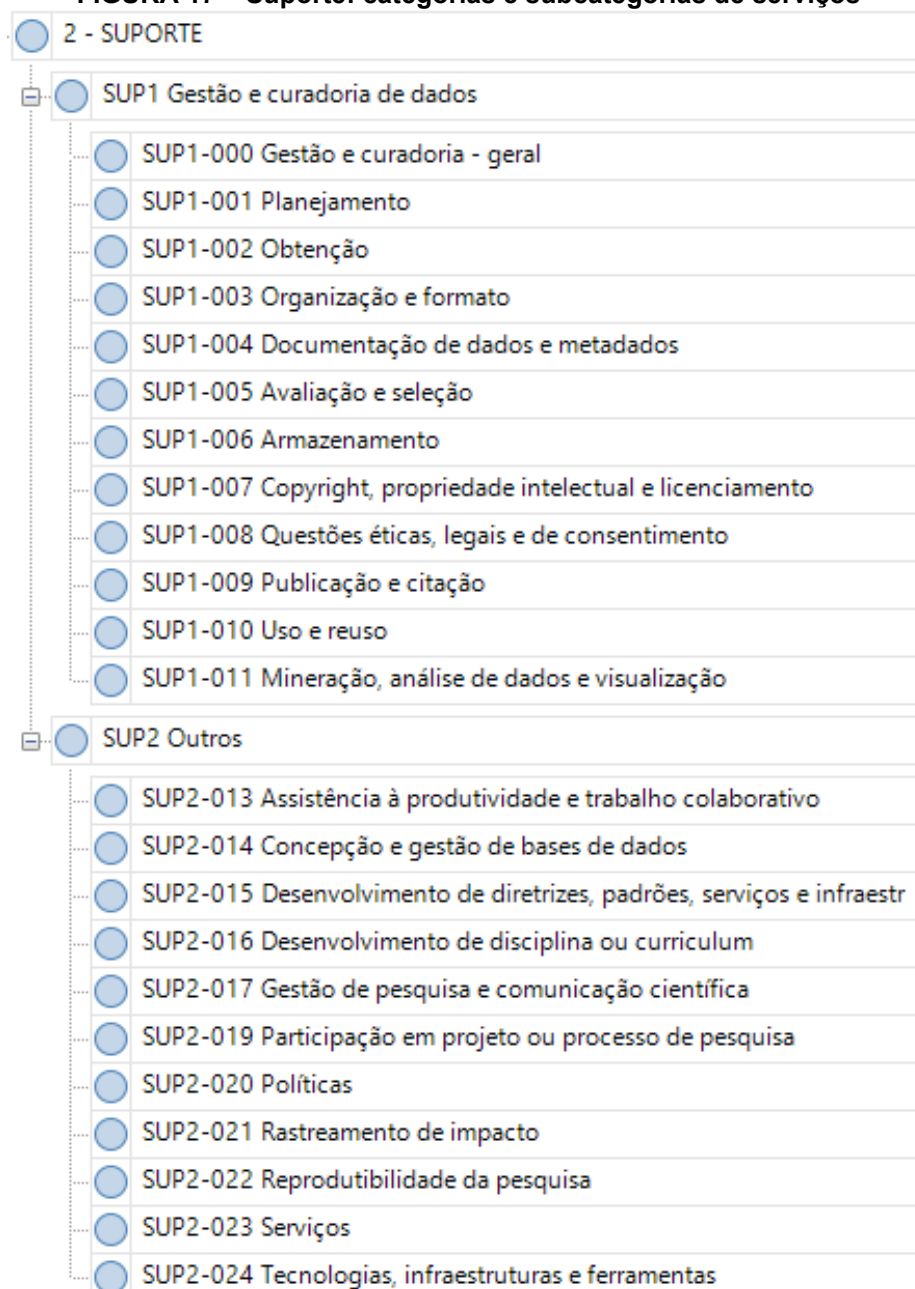
FIGURA 15 - Formação dos nomes das subcategorias

Fonte: Elaborado pela autora.

É importante frisar que as subcategorias ‘Gestão e curadoria de dados’ e ‘Outros’ estão presentes em todas as categorias principais, mesmo que de forma parcial, conforme pode ser observado nas FIG. 16 e 17.

FIGURA 16 – Infraestrutura: categorias e subcategorias de serviços

Fonte: Elaborado pela autora.

FIGURA 17 – Suporte: categorias e subcategorias de serviços

Fonte: Elaborado pela autora.

4.1.1. Conteúdo das subcategorias

A subcategoria 'Gestão e curadoria de dados', independentemente da categoria principal (Infraestrutura, Suporte e Educacional), é formada por serviços que abordam os seguintes assuntos (Quadro 21):

QUADRO 21 – Gestão e curadoria de dados: assuntos presentes nas subcategorias

000 Gestão e curadoria – geral
Assuntos gerais: Gestão, inclusive de dados ativos. Curadoria. Preservação. Dados de pesquisa. Ciclo de vida dos dados. <i>Data stewardship</i> . Manutenção de dados. Melhores práticas de gestão e de compartilhamento. Estratégias para gestão e disseminação de dados. <i>Data Science</i> , <i>eScience</i> , <i>eResearch</i> .
001 Planejamento
Plano de Gestão de Dados (PGD). Planejamento da criação, armazenamento e preservação de dados. Estratégias de preservação. <i>Design</i> de pesquisa. Financiamento. Conformidade com os requisitos dos financiadores e editores.
002 Obtenção
Criação, coleta ou captura de dados. Digitalização. Aquisição de dados. Licenciamento/assinatura de dados. Inclui a busca, localização ou descoberta de dados. Catálogo de dados.
003 Organização e formato
Organização de arquivos e pastas (nomes de arquivos e estruturação de pastas). Uso de formato recomendados de dados e padrões intercambiáveis. Padronização. Formatos de preservação. Conversão de formato. Migração. Garantia de qualidade dos dados (entrada de dados, digitalização, transcrição e verificação). Controle de versão. Validação. Empacotamento. Autenticidade, confiabilidade, compreensibilidade dos dados. Limpeza de dados. Preparação de dados.
004 Documentação de dados e metadados
Anotação. Descrição (informações que expliquem como os dados foram criados, coletados, adquiridos ou digitalizados, o que eles significam, seu conteúdo e estrutura, e quaisquer manipulações ocorridas). Padrões para descrição de dados. Atribuição de metadados padronizados e estruturados (para descrição e preservação). Geração automática de metadados. Dicionário de dados. Descrição da metodologia.
005 Avaliação e seleção
Avaliação e seleção de dados para preservação (depósito em repositórios). Reavaliação de dados ou conjunto de dados que devem ser preservados.
006 Armazenamento
Armazenamento em curto e longo prazo. Arquivamento. Repositório de dados. Serviços de repositório. Identificação de repositórios. <i>Backup</i> e segurança (prevenção de perdas físicas – acidental, falhas de <i>hardware</i> e <i>software</i> , desastres, erros humanos, corrupção de dados, acesso não autorizado). Criptografia. Espelhamento e replicação. Eliminação com segurança. Integridade dos dados. Compartilhamento e depósito. Disseminação. Mover dados. Transferência. Ingestão. Redundância de dados.
007 Copyright, propriedade intelectual e licenciamento
Proteção de dados (dos direitos autorais, da propriedade intelectual). Direito autoral (<i>copyright</i>). Direitos de propriedade intelectual. Autoria e propriedade dos dados. Licenciamento para dados. Condições para obtenção de licença. Controle de acesso. Embargo.
008 Questões éticas, legais e de consentimento
Privacidade e confidencialidade de dados. Dados sensíveis. Anonimização. Obrigações éticas com a pesquisa. Consentimento para compartilhamento de dados.
009 Publicação e citação
Publicação de dados. Citação de dados.
010 Uso e reuso
Garantia de acesso. Acesso permanente e acesso aberto. Identificadores persistentes (DOI, ORCID). Ligação de conjunto de dados aos documentos. Registro de dados. Uso e reuso. Disponibilidade. Garantia de <i>links</i> apropriados de dados suplementares para artigos em repositórios institucionais.
011 Mineração, análise e visualização
Mineração. Análise. Visualização. Processamento. Integração. Criação de gráficos e tabelas para dados numéricos. Computação de alto desempenho – análise avançada, simulação, modelagem, inteligência artificial.

Fonte: Elaborado pela autora.

A subcategoria 'Outros' é formada por serviços que abordam os seguintes assuntos (Quadro 22):

QUADRO 22 – Outros: assuntos presentes nas subcategorias

012 Análise de redes sociais
Ferramentas e métodos para análise de redes sociais.
013 Assistência à produtividade e trabalho colaborativo
Ferramentas e sistemas para assistência à produtividade e trabalho colaborativo.
014 Concepção e gestão de bases de dados
Concepção e gestão de bases de dados.
015 Desenvolvimento de diretrizes, padrões, serviços e infraestruturas
Suporte (informação, orientação, aconselhamento e consultoria) relacionado ao desenvolvimento de políticas, diretrizes, padrões, serviços e infraestruturas.
016 Desenvolvimento de disciplina ou <i>curriculum</i>
Participação no desenvolvimento de disciplina ou currículo de disciplina voltado para a gestão e curadoria de dados nas escolas de Biblioteconomia. Inclui o ensino de gestão e curadoria de dados.
017 Gestão de pesquisa e comunicação científica
Fluxos de pesquisa. Gestão da informação de pesquisas conduzidas na instituição. Inclui a <i>scholarly communication</i> - resultados das pesquisas são avaliados, disseminados para a comunidade acadêmica e preservados para uso futuro em meios formais e informais de comunicação.
018 Identificação das percepções, práticas e necessidades
Eventos que abordam a importância de identificação das necessidades dos pesquisadores em relação à gestão de dados.
019 Participação em projeto ou processo de pesquisa
Participação da biblioteca em projeto ou processo de pesquisa, atuando de forma mais próxima ao pesquisador.
020 Políticas
Política institucional para a gestão de dados que inclui a criação de um plano de gestão de dados (PGD), o compartilhamento, acesso e suporte aos dados, submissão de dados em repositórios, licenciamento para reutilização, período de embargo, normas disciplinares para garantir a integridade da pesquisa e o cumprimento dos requisitos das agências financiadoras. Desenvolvimento de políticas.
021 Rastreamento de impacto
Impacto do uso de dados. Análise métrica sobre o uso. Impacto do uso e reuso de dados. Impacto do compartilhamento.
022 Reprodutibilidade da pesquisa
Questões de reprodutibilidade da pesquisa relacionadas à gestão adequada dos dados de pesquisa, tornando-os amplamente disponíveis.
023 Serviços
<i>Links</i> para repositórios, <i>data centers</i> , ferramentas, serviços de gestão de dados. Orientação ou treinamentos - oficinas, cursos, treinamentos - relacionados aos serviços oferecidos pelas bibliotecas.
024 Tecnologias, infraestruturas e ferramentas
Outras tecnologias, infraestruturas e ferramentas não relacionadas diretamente à gestão e curadoria de dados, mas que são oferecidas aos pesquisadores. Inclui: <i>Survey tools</i> . Programação em Python. LaTeX. <i>Bash scripting</i> . Unix. Controle de versão com GIT. <i>Online survey</i> . Redes. Dispositivos móveis.

Fonte: Elaborado pela autora.

4.1.1.1. Categoria Infraestrutura

A categoria 'Infraestrutura' inclui serviços relacionados à disponibilização de ferramentas, métodos, *softwares* e plataformas para a gestão e curadoria de dados de pesquisa. Inclui os repositórios de dados, além de outras infraestruturas como políticas institucionais, ferramentas para análise de redes sociais, assistência à produtividade e trabalho colaborativo, infraestruturas de redes, dentre outras.

As infraestruturas são classificadas em gestão e curadoria de dados – INF1 Gestão e curadoria de dados – e outras infraestruturas – INF2 Outros. A disposição das subcategorias constantes da categoria '1 – Infraestrutura' é apresentada na FIG. 18.

FIGURA 18 – Infraestrutura: visão geral das subcategorias



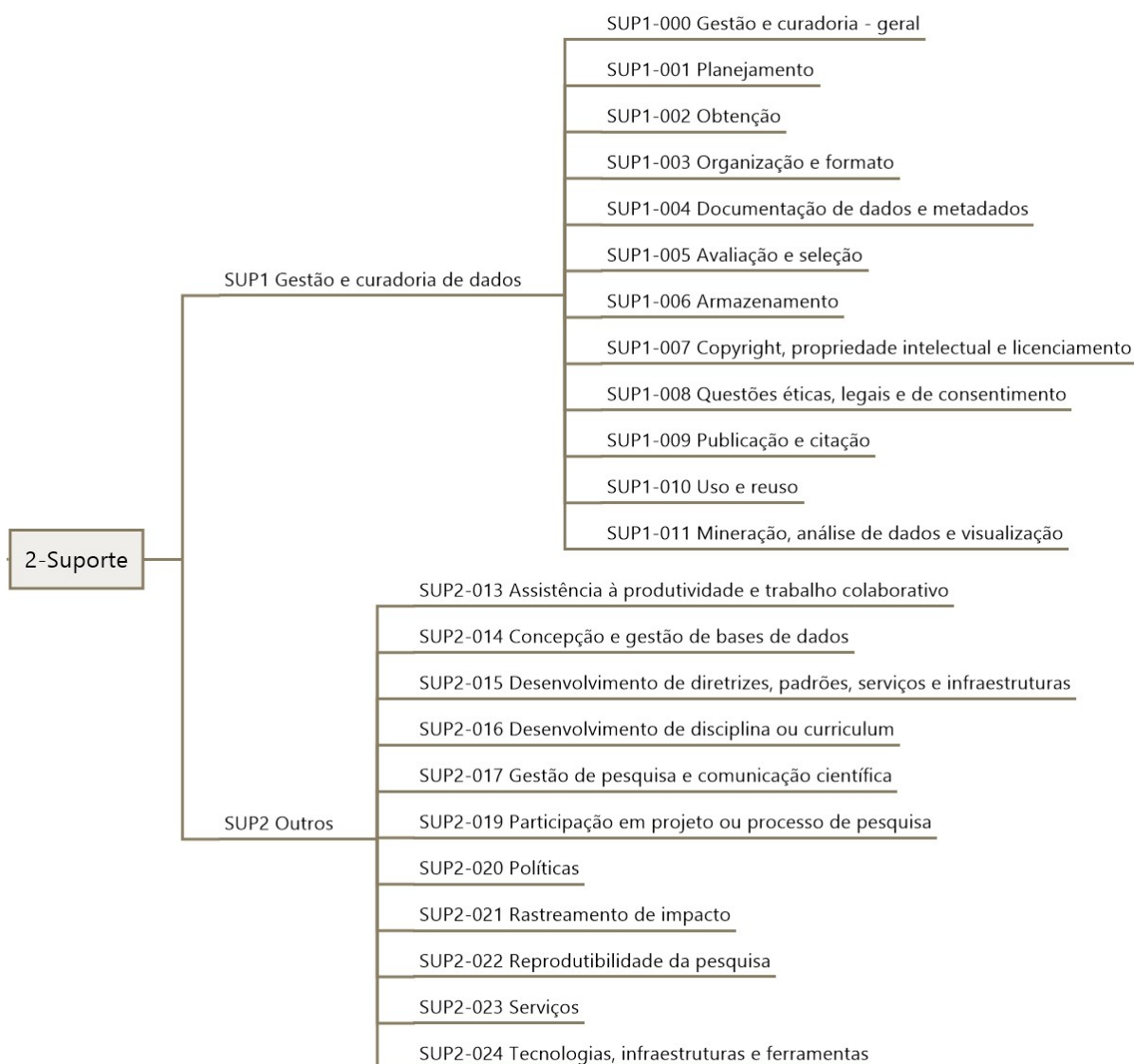
Fonte: Elaborado pela autora.

4.1.1.2. Categoria Suporte

A categoria 'Suporte' é formada pelos serviços técnicos (1), tipo assistência técnica, onde a própria biblioteca executa uma série de atividades, como: documentação de dados e metadados; depósito de dados; preservação; garantia de acesso; publicação; dentre outros. Inclui também atividades de suporte (2) como informação, orientação, aconselhamento e consultoria, de modo presencial ou *online*, sobre os mais variados assuntos referentes à gestão e curadoria, como a elaboração de planos de gestão de dados, padrões de metadados, acesso aberto aos dados, melhores práticas de pesquisa, políticas de dados, serviços oferecidos e outros.

Esses serviços podem ser classificados em suporte voltado para a gestão e curadoria de dados – SUP1 Gestão e curadoria de dados - e outros tipos de suporte – SUP2 Outros. A FIG. 19 apresenta a estruturação da categoria ‘2 – Suporte’.

FIGURA 19 – Suporte: visão geral das subcategorias



Fonte: Elaborado pela autora.

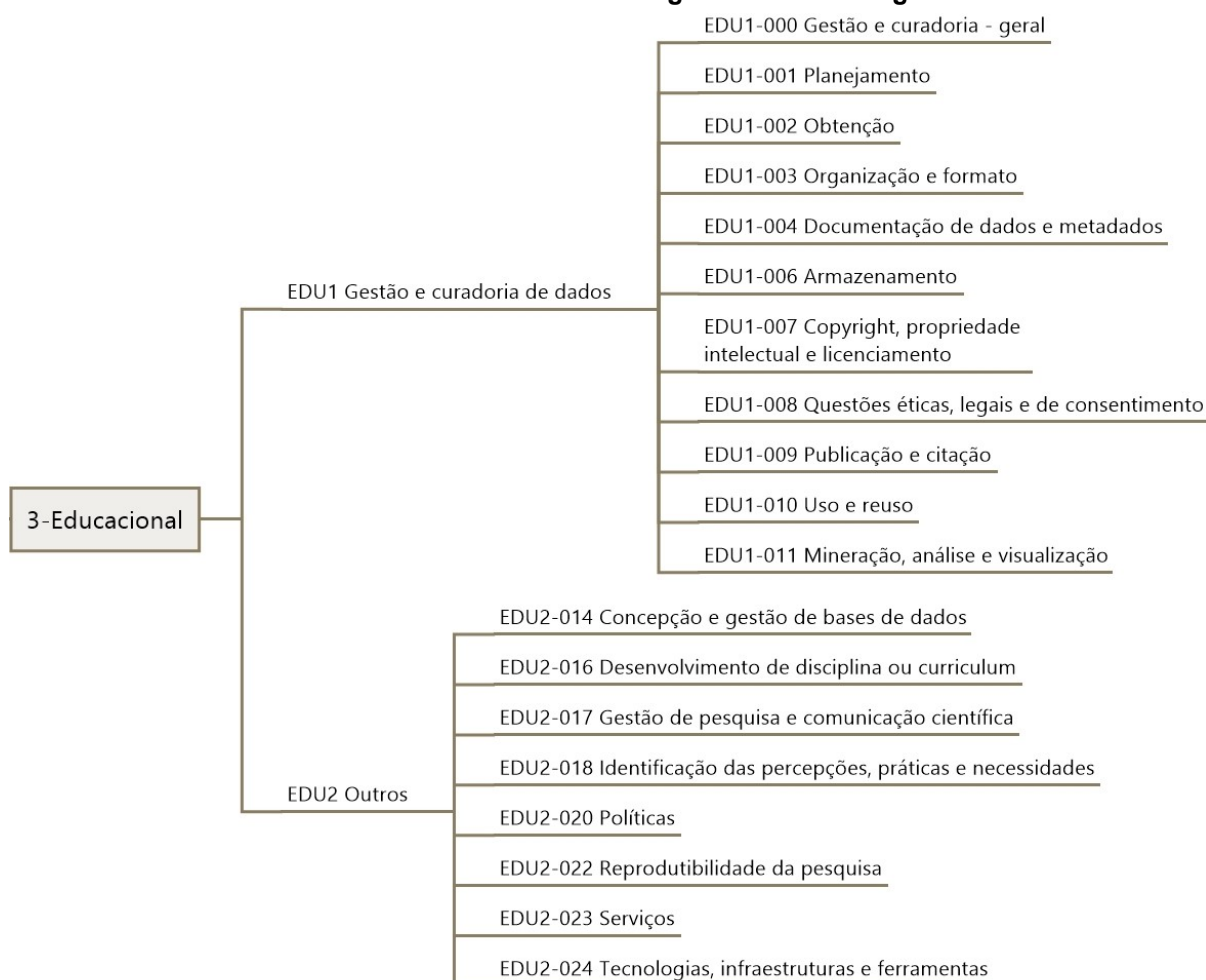
4.1.1.3. Categoria Educacional

A categoria ‘Educacional’ abrange os serviços educacionais voltados para a formação em gestão e curadoria de dados realizados pelas bibliotecas como *advocacy*, conscientização, incentivo em questões de gestão e curadoria de dados. Inclui a conscientização sobre os serviços oferecidos pelas bibliotecas; conferências, seminários, simpósios, palestras e painéis relacionados à gestão e curadoria de dados.

Pertencem também à categoria outros eventos relacionados à identificação das necessidades dos pesquisadores, políticas, reprodutibilidade da pesquisa e serviços

oferecidos; elaboração e disponibilização de guias, *templates*, tutoriais ou *help* relacionados à gestão e curadoria de dados e questões relacionadas às políticas e tecnologias, infraestruturas e ferramentas; treinamentos, oficinas, cursos, aulas relacionadas à gestão e curadoria de dados, bem como outros tipos de treinamentos como àqueles relacionados às questões políticas, serviços, tecnologias, infraestruturas e ferramentas oferecidas. Os serviços educacionais são classificados em gestão e curadoria de dados – EDU1 Gestão e curadoria de dados - e outros tipos de serviços – EDU2 Outros. A estruturação desta categoria é apresentada na FIG. 20.

FIGURA 20 – Educacional: visão geral das subcategorias



Fonte: Elaborado pela autora.

Alguns trechos retirados dos artigos ilustram a criação das três categorias principais, suas subcategorias e subdivisões:

a) categoria Infraestrutura;

INF1-001 Planejamento - **Data management plan** templates and tools (BERMAN, 2017c, p. 13, grifo nosso).

INF1-002 Obtenção - [...] providing tools for dataset discovery [...] (SWANSON; RINEHART, 2016, p. 98).

INF1-011 Mineração, análise e visualização - [...] providing tools for data mining and visualization [...] (TENOPIR et al., 2017, p. 27).

INF2-013 Assistência à produtividade e trabalho colaborativo - [...] include web-based tools for collaborative writing, electronic note-taking [...] (JOHNSON; BUTLER; JOHNSTON, 2012, p. 7).

b) categoria Suporte;

SUP1-003 Organização e formato - [...] libraries provided information about how to properly organize data files (YOON; SCHULTZ, 2017, p. 926).

SUP2-015 Desenvolvimento de diretrizes, padrões, serviços e infraestruturas - [...] advise and help develop publishing standards and systems (GOLDb, 2007, não paginado).

SUP2-020 Políticas - Assisting in developing policies about data management and preservation (BROWN; WOLSKI; RICHARDSON, 2015, p. 226).

c) categoria Educacional.

EDU1-003 Organização e formato - A workshop on types, formats, and stages of data [...] (PETERS; VAUGHN, 2014, p. 95).

EDU1-009 Publicação e citação de dados - Some programs also briefly touched upon topics relevant to data management (such as publishing your data) during data-related classes or workshops (YOON; SCHULTZ, 2017, p. 928).

EDU2-023 Serviços - [...] **training sessions** at faculty and graduate students, as well as meeting with the Office of Research and Sponsored Programs **to explain the new requirement and RDS' available support services** (RABOIN; REZNIK-ZELLEN; SALO, 2012, p. 136, grifo nosso).

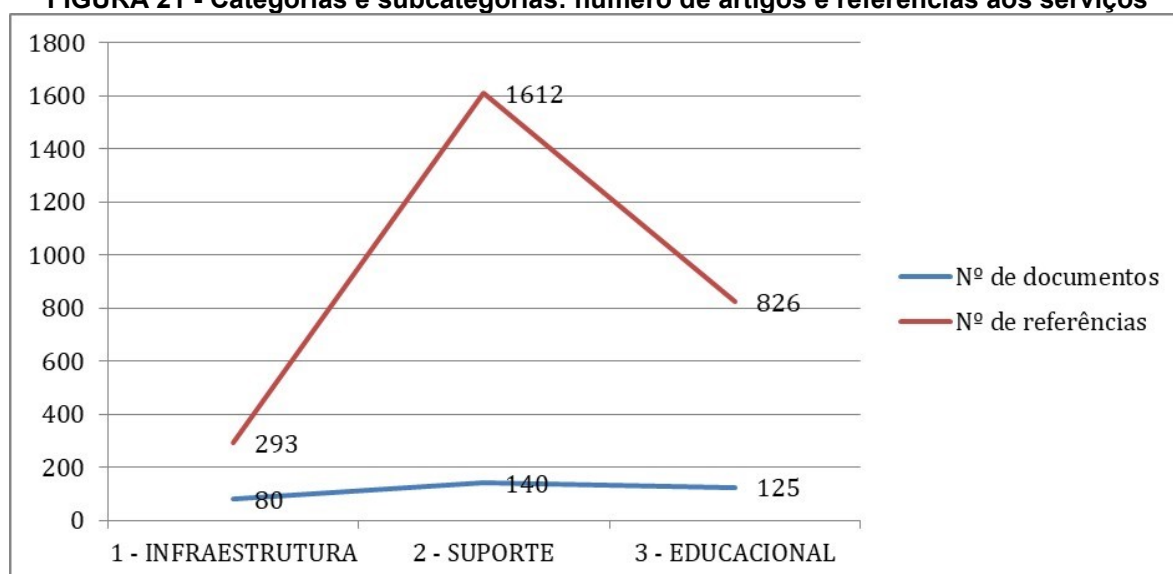
Fragmentos dos textos exemplificando a criação e inclusão dos serviços em todas categorias, subcategorias e subdivisões presentes na taxonomia podem ser consultados nos Apêndices E, F e G.

A estruturação completa da taxonomia de serviços e quantificação de referências aos serviços incluídas em cada uma das categorias, subcategorias e subdivisões podem ser consultados nos Apêndices H e I.

4.2. TAXONOMIA DE SERVIÇOS: ANÁLISE DESCRITIVA E COMPILAÇÃO DE SSGDPs

Os serviços categorizados como 'Suporte' foram os mais presentes. Constataram-se 1612 citações, o que corresponde a 59,03% das referências aos serviços encontradas nos 167 documentos (165 artigos + duas entrevistas) utilizados para a construção da taxonomia. Em seguida aparecem os serviços da categoria 'Educação', com 826 referências, representando 30,24% das citações. Por último, os serviços da categoria 'Infraestrutura', com 293 referências, representaram 10,73% do total das citações aos serviços encontradas, conforme pode ser observado no gráfico a seguir (FIG. 21).

FIGURA 21 - Categorias e subcategorias: número de artigos e referências aos serviços



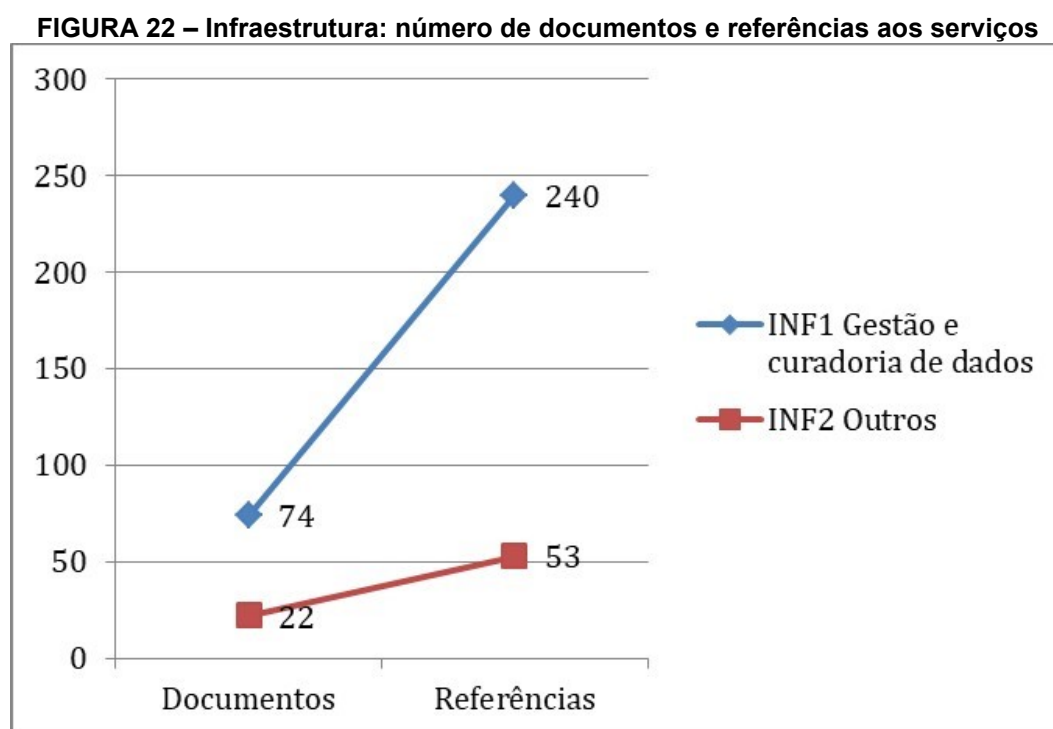
Fonte: Dados de pesquisa (2019).

A seguir é apresentada uma análise descritiva de cada uma das categorias e subcategorias de serviços.

4.2.1. Categoria Infraestrutura

Para esta categoria, as infraestruturas tecnológicas foram consideradas serviços que as bibliotecas e/ou instituições podem oferecer. A categoria 'Infraestrutura' é formada pela disponibilização de ferramentas, métodos, *softwares* e plataformas para gestão e curadoria de dados. Inclui também os repositórios de dados e outros tipos de infraestruturas, como as políticas institucionais, importantes para o funcionamento dos SSGDPs nas instituições. Também fazem parte da categoria as ferramentas para análise de redes sociais, ferramentas de assistência à produtividade e trabalho colaborativo, infraestruturas de rede, entre outras.

Para a formação da categoria '1 – Infraestrutura' foram utilizados 80 documentos. Deles, 74 faziam referência aos serviços da subcategoria 'INF1 Gestão e curadoria de dados'. Outros 22 documentos faziam referências aos serviços da subcategoria 'INF2 Outros'. Foram encontradas, ainda, 240 menções aos serviços da subcategoria 'INF1 Gestão e curadoria de dados' e 53 menções aos serviços da categoria 'INF2 Outros', conforme pode ser observado na FIG. 22.



Fonte: Dados de pesquisa (2019).

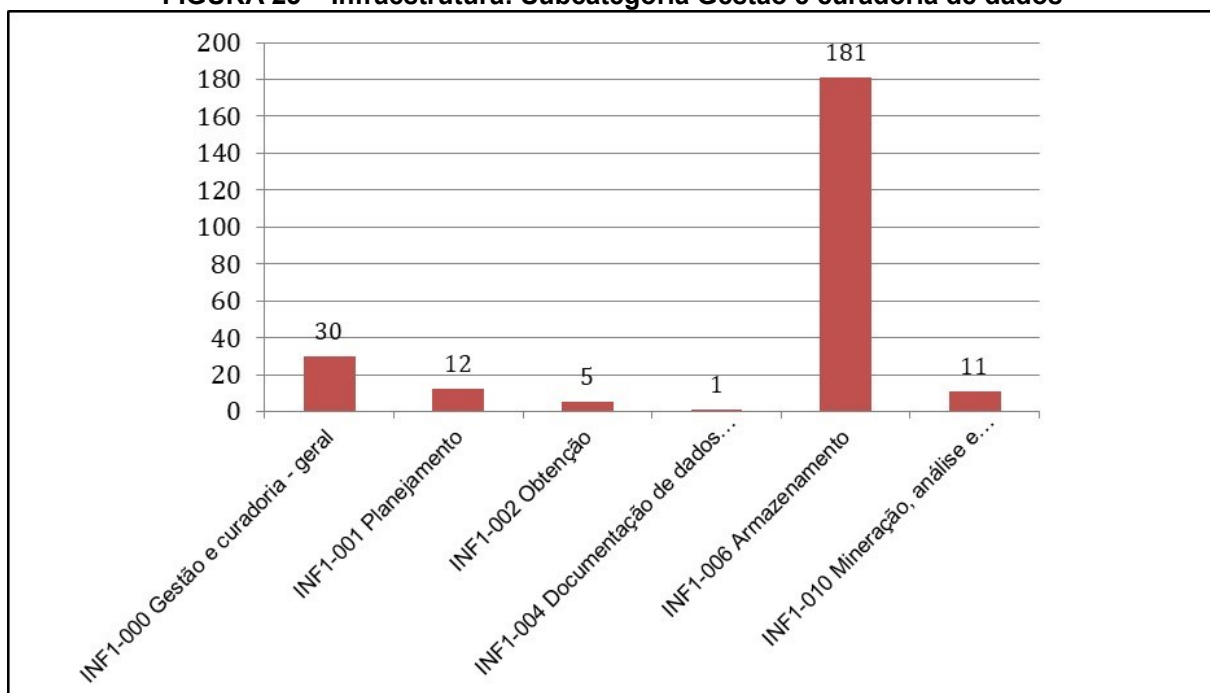
Assim, das 293 referências aos serviços da categoria 'Infraestrutura', observamos que 240 delas, representando 81,91% dos serviços referenciados na categoria e, 8,79% do total de serviços encontrados, dizem respeito à gestão e curadoria de dados – INF1 Gestão e curadoria de dados (TAB. 1).

TABELA 1 – Infraestrutura: frequência de referências aos serviços

1- INFRAESTRUTURA		N. de ref. (abs.)	% na categoria	% do total
INF1 Gestão e curadoria de dados	INF1-000 Gestão e curadoria - geral	30	10,24	1,10
	INF1-001 Planejamento	12	4,10	0,44
	INF1-002 Obtenção	5	1,71	0,18
	INF1-004 Documentação de dados e metadados	1	0,34	0,04
	INF1-006 Armazenamento	181	61,77	6,63
	INF1-011 Mineração, análise de dados e visualização	11	3,75	0,40
INF1 Gestão e curadoria de dados		240	81,91	8,79
INF2 Outros	INF2-013 Assistência à produtividade e trabalho colaborativo	13	4,44	0,48
	INF2-012 Análise de redes sociais	1	0,34	0,04
	INF2-017 Gestão de pesquisa e comunicação científica	4	1,37	0,15
	INF2-020 Políticas	29	9,90	1,06
	INF2-024 Tecnologias, infraestruturas e ferramentas	6	2,05	0,22
INF2 Outros		53	18,09	1,94
Total de referências na categoria 1 - Infraestrutura		293	100,00	10,73
Total geral de referências		2731		

Fonte: Dados de pesquisa (2019).

Na subcategoria ‘INF1 Gestão e curadoria de dados’, destacam-se as infraestruturas para armazenamento de dados – INF1-006 Armazenamento -, com 181 referências encontradas, perfazendo um total de 61,77% das referências aos serviços da categoria ‘1 – Infraestrutura’, e 6,63% do total de referências aos diversos serviços encontrados nas três categorias (TAB. 1 e FIG. 23).

FIGURA 23 – Infraestrutura: Subcategoria Gestão e curadoria de dados

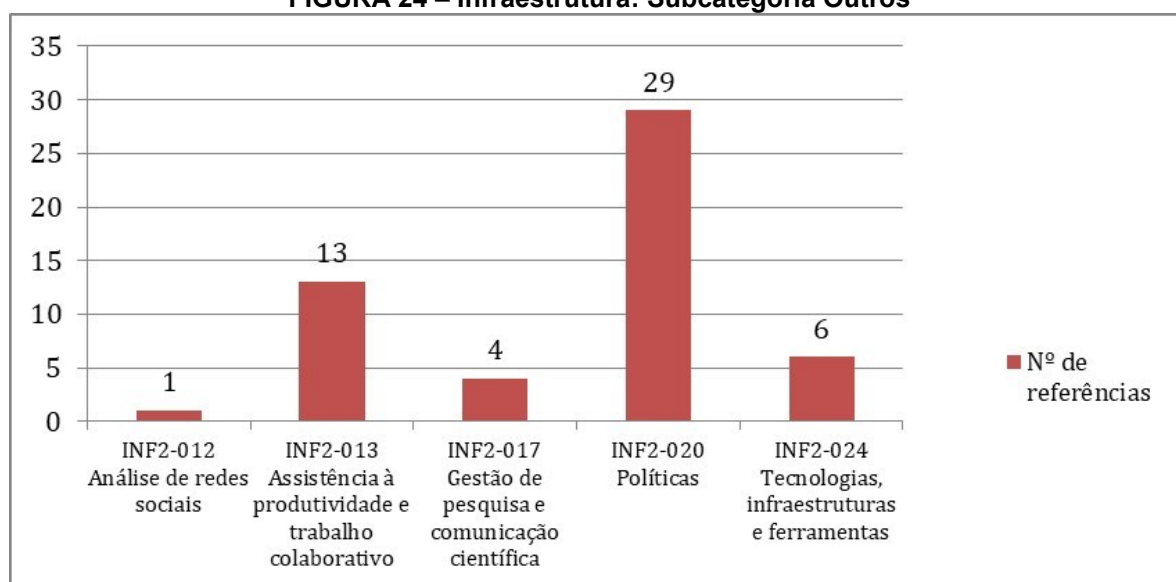
Fonte: Dados de pesquisa (2019).

Os serviços da categoria 'Infraestrutura', subcategoria 'Gestão e curadoria de dados', subdivisão 'INF1-006 Armazenamento' são o fornecimento ou desenvolvimento de:

- a) ferramenta para publicação de dados ou conjunto de dados;
- b) repositórios de dados, específicos ou multidisciplinares – infraestrutura que suporte o arquivamento, compartilhamento, preservação e reuso de dados;
- c) serviços de repositórios como estatísticas e relatórios, preservação de dados e assistência com propriedade intelectual e direitos autorais;
- d) plataforma de troca e compartilhamento de dados;
- e) ferramentas ou serviços de *backup* de dados.

Na subcategoria 'INF2 Outros' foram encontradas 53 alusões às infraestruturas, perfazendo um total de 18,09% das referências aos serviços da categoria 'Infraestrutura' e 1,94% do total de referências marcadas, com destaque para a infraestrutura política – INF2-020 Políticas - que apresenta 29 menções, ou seja, 9,90% das referências aos serviços da categoria 'Infraestrutura' e 1,06% do total de serviços mencionados, seguido pelos serviços categorizados como 'INF2-013 Assistência à produtividade e trabalho colaborativo' com 13 referências, ou 4,44% dos serviços da categoria 'Infraestrutura' e 0,48% do total de serviços (TAB. 1 e FIG. 24).

FIGURA 24 – Infraestrutura: Subcategoria Outros



Fonte: Dados de pesquisa (2019).

A categoria 'Infraestrutura', subcategoria 'Outros', 'INF2-020 Políticas' compreende a provisão de:

- a) políticas institucionais para a gestão de dados;

- b) políticas para acesso aos dados submetidos no repositório, licenciamento para reutilização de dados e período de embargo.

Já a subcategoria 'Outros', subdivisão 'INF2-013 Assistência à produtividade e trabalho colaborativo' compreende o fornecimento de:

- a) ferramenta para anotações eletrônicas;
- b) ferramenta para compartilhamento de citações;
- c) plataforma para pesquisa colaborativa (compartilhamento e produção de conhecimento de forma colaborativa, escrita colaborativa).

Embora menos referenciada, a categoria 'Infraestrutura' abrange importantes alicerces para os SSGDPs e para gestão adequada dos dados produzidos nas instituições, como os repositórios de dados, incluídos na subcategoria 'INF1-006 Armazenamento', e as políticas institucionais contidas na subcategoria 'INF2-020 Políticas'.

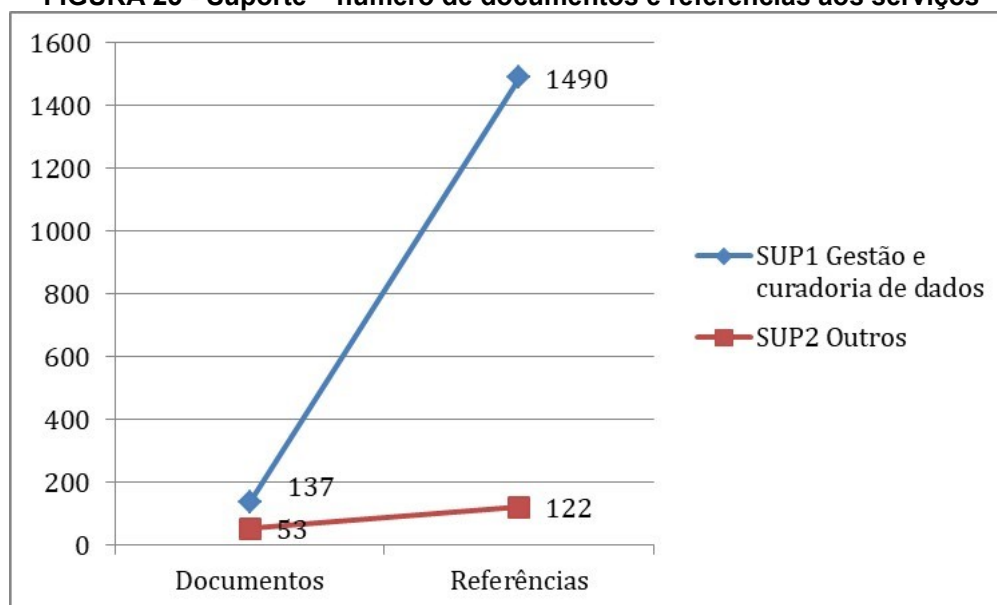
Os repositórios de dados são considerados parte fundamental da infraestrutura, conforme defendem autores como Sayão e Sales (2016), Tenopir *et al.* (2013, 2015, 2017), Rodrigues *et al.* (2010) e outros. As políticas institucionais de gestão, curadoria, compartilhamento e preservação de dados também são consideradas indispensáveis (LYON, 2007; SHEARER, 2015) aos serviços de suporte e à efetiva gestão de dados.

A lista com todos os serviços da categoria '1 – Infraestrutura' pode ser consultada no Apêndice J.

4.2.2. Categoria Suporte

Na categoria 'Suporte', formada pelos serviços técnicos (1) realizados pelas bibliotecas, tipo de assistência técnica, onde a própria biblioteca executa uma série de atividades como: documentação de dados e metadados; depósito de dados; disponibilização de identificadores únicos; atividades de suporte (2) propriamente ditas como a informação, orientação, aconselhamento e consultoria, de modo presencial ou *online*. O suporte pode versar sobre os mais variados assuntos referentes à gestão e curadoria de dados como: elaboração de planos de gestão de dados; padrões de metadados; acesso aberto aos dados; melhores práticas de pesquisa; políticas de dados; serviços oferecidos e outros.

Nesta categoria, 137 documentos fizeram 1490 menções aos serviços da subcategoria 'SUP1 Gestão e curadoria de dados', enquanto 53 mencionaram 122 vezes os serviços da subcategoria 'SUP 2 Outros', como observado na FIG. 25.

FIGURA 25 - Suporte – número de documentos e referências aos serviços

Fonte: Dados de pesquisa (2019)

Das 1612 referências aos serviços da categoria 'Suporte', 1490, ou seja, 92,43% dos serviços marcados nesta categoria e 54,56% do total de serviços marcados estão presentes na subcategoria 'SUP1 Gestão e curadoria de dados'. Outras 122 referências, representando 7,57% dos serviços da categoria 'Suporte' e 4,47% do total de referências encontradas estão presentes na subcategoria 'SUP2 Outros' (TAB. 2).

TABELA 2 – Suporte: frequência de referências aos serviços

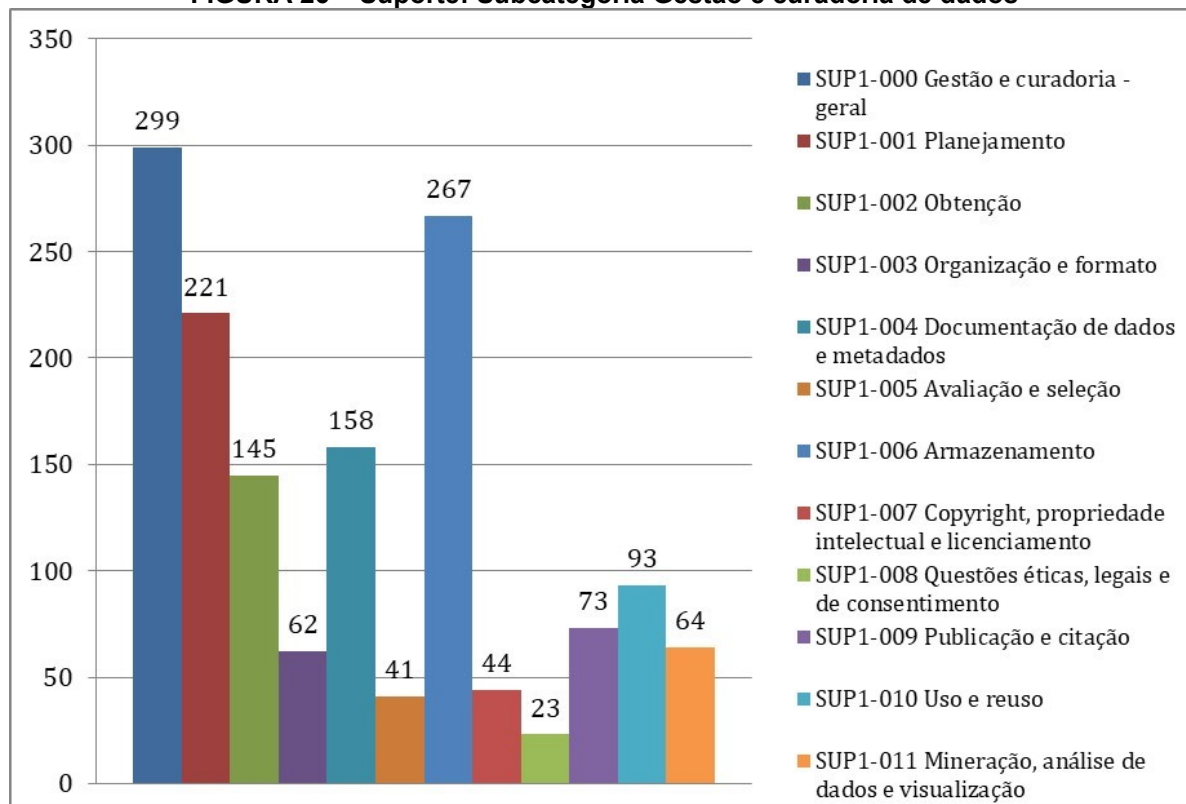
		(continua)		
2 – SUPORTE		N. de ref. (abs.)	% na categoria	% do total
SUP1 Gestão e curadoria de dados	SUP1-000 Gestão e curadoria – geral	299	18,55	10,95
	SUP1-001 Planejamento	221	13,71	8,09
	SUP1-002 Obtenção	145	9,00	5,31
	SUP1-003 Organização e formato	62	3,85	2,27
	SUP1-004 Documentação de dados e metadados	158	9,80	5,79
	SUP1-005 Avaliação e seleção	41	2,54	1,50
	SUP1-006 Armazenamento	267	16,56	9,78
	SUP1-007 <i>Copyright</i> , propriedade intelectual e licenciamento	44	2,73	1,61
	SUP1-008 Questões éticas, legais e de consentimento	23	1,43	0,84
	SUP1-009 Publicação e citação	73	4,53	2,67
	SUP1-010 Uso e reuso	93	5,77	3,41
	SUP1-011 Mineração, análise de dados e visualização	64	3,97	2,34
SUP1 Gestão e curadoria de dados		1490	92,43	54,56
SUP2 Outros	SUP2-013 Assistência à produtividade e trabalho colaborativo	1	0,06	0,04
	SUP2-014 Concepção e gestão de bases de dados	2	0,12	0,07
	SUP2-015 Desenvolvimento de diretrizes, padrões, serviços e infraestruturas	7	0,43	0,26

TABELA 2 – Suporte: frequência de referências aos serviços

(conclusão)			
2 – SUPORTE	N. de ref. (abs.)	% na categoria	% do total
SUP2-016 Desenvolvimento de disciplina ou <i>curriculum</i>	3	0,19	0,11
SUP2-017 Gestão de pesquisa e comunicação científica	5	0,31	0,18
SUP2-019 Participação em projeto ou processo de pesquisa	24	1,49	0,88
SUP2-020 Políticas	12	0,74	0,44
SUP2-021 Rastreamento de impacto	19	1,18	0,70
SUP2-022 Reprodutibilidade da pesquisa	3	0,19	0,11
SUP2-023 Serviços	11	0,68	0,40
SUP2-024 Tecnologias, infraestruturas e ferramentas	35	2,17	1,28
SUP2 Outros	122	7,57	4,47
Total de referências na categoria 2 - Suporte	1612	100,00	59,03
Total geral de referências		2731	

Fonte: Dados de pesquisa (2019).

O suporte para armazenamento – SUP1-006 Armazenamento – com 267 citações (16,56% das referências marcadas na categoria) e suporte às atividades de planejamento – SUP1-001 Planejamento - com 221 citações (13,71% das referências marcadas na categoria) são considerados os serviços de suporte mais presentes nos artigos avaliados (TAB. 2 e FIG. 26), embora a subdivisão ‘SUP1-000 Gestão e curadoria – geral’ - apresente 299 (18,55%) das referências marcadas na subcategoria.

FIGURA 26 – Suporte: Subcategoria Gestão e curadoria de dados

Fonte: Dados de pesquisa (2019).

A subcategoria ‘SUP1-000 Gestão e curadoria – geral’ não foi considerada como a mais frequente. Nela foram incluídos os serviços da subcategoria ‘SUP1 Gestão e curadoria de dados’ que não foram especificados, tendo sido apresentados de forma generalista, vaga ou ampla, como pode ser observado nos trechos apresentados na subdivisão ‘SUP1-000 Gestão e curadoria – geral’ apresentada a seguir.

The Library will provide data services for acquisition, **management, and preservation of massive amounts of data** [...] (CLAIBOURN, 2015, p. 104, grifo nosso).
 [...] **data management guidance during research**” (GOLDMAN; KAFEL; MARTIN, 2015, p. 2).
 [...] **support the data curation** [...] (GOLDMAN; KAFEL; MARTIN, 2015, p. 2)
 [...] **assisting researchers with managing** [...] (CORRALL; KENNAN; AFZAL, 2013, p. 655, grifo nosso).

Excluindo os serviços da subdivisão ‘SUP1-000 Gestão e curadoria – geral’, a subdivisão ‘SUP1-006 Armazenamento’ passa a corresponder a 20,34% das referências da subcategoria. Já o suporte às atividades de planejamento – SUP1-001 Planejamento - passa a responder por 16,83% das referências marcadas na subcategoria, conforme apresentado na TAB. 3.

TABELA 3 – Suporte: total de referências na categoria

(continua)

2 – SUPORTE		N. de ref. (abs.)	% na categoria	%, exceto SUP1-000
SUP1 Gestão e curadoria de dados	SUP1-000 Gestão e curadoria – geral	299	18,55	-
	SUP1-001 Planejamento	221	13,71	16,83
	SUP1-002 Obtenção	145	9,00	11,04
	SUP1-003 Organização e formato	62	3,85	4,72
	SUP1-004 Documentação de dados e metadados	158	9,80	12,03
	SUP1-005 Avaliação e seleção	41	2,54	3,12
	SUP1-006 Armazenamento	267	16,56	20,34
	SUP1-007 <i>Copyright</i> , propriedade intelectual e licenciamento	44	2,73	3,35
	SUP1-008 Questões éticas, legais e de consentimento	23	1,43	1,75
	SUP1-009 Publicação e citação	73	4,53	5,56
	SUP1-010 Uso e reuso	93	5,77	7,08
	SUP1-011 Mineração, análise de dados e visualização	64	3,97	4,87
SUP2 Outros	SUP2-013 Assistência à produtividade e trabalho colaborativo	1	0,06	0,08
	SUP2-014 Concepção e gestão de bases de dados	2	0,12	0,15
	SUP2-015 Desenvolvimento de diretrizes, padrões, serviços e infraestruturas	7	0,43	0,53
	SUP2-016 Desenvolvimento de disciplina ou <i>curriculum</i>	3	0,19	0,23
	SUP2-017 Gestão de pesquisa e comunicação científica	5	0,31	0,38
	SUP2-019 Participação em projeto ou processo de pesquisa	24	1,49	1,83
	SUP2-020 Políticas	12	0,74	0,91
	SUP2-021 Rastreamento de impacto	19	1,18	1,45
	SUP2-022 Reprodutibilidade da pesquisa	3	0,19	0,23

TABELA 3 – Suporte: total de referências na categoria

		(conclusão)		
2 – SUPORTE		N. de ref. (abs.)	% na categoria	%, exceto SUP1-000
SUP2 Outros	SUP2-023 Serviços	11	0,68	0,84
	SUP2-024 Tecnologias, infraestruturas e ferramentas	35	2,17	2,67
Total de referências na categoria		1612	100,00	-
Total de referência na categoria, exceto SUP1-000		1313	-	100,00

Fonte: Dados de pesquisa (2019).

A categoria ‘Suporte’, subcategoria ‘Gestão e curadoria de dados’, ‘SUP1 006 Armazenamento’ é constituída por serviços de fornecimento de suporte, orientação, assistência, aconselhamento, recomendações e consultoria para:

- a) armazenamento/arquivamento/depósito de dados ou conjunto de dados em repositório;
- b) implementação de repositório de dados;
- c) uso de repositórios de dados;
- d) depósito de dados em repositório internos e externos;
- e) identificação de repositórios apropriados para depositar dados;
- f) compartilhamento de dados;
- g) determinação do melhor repositório para um dado conjunto de dados;
- h) suporte técnico para repositórios de dados.

A subdivisão ‘SUP1 001 Planejamento’ é formada por serviços como o fornecimento de suporte, orientação, assistência, aconselhamento, recomendações e consultoria para:

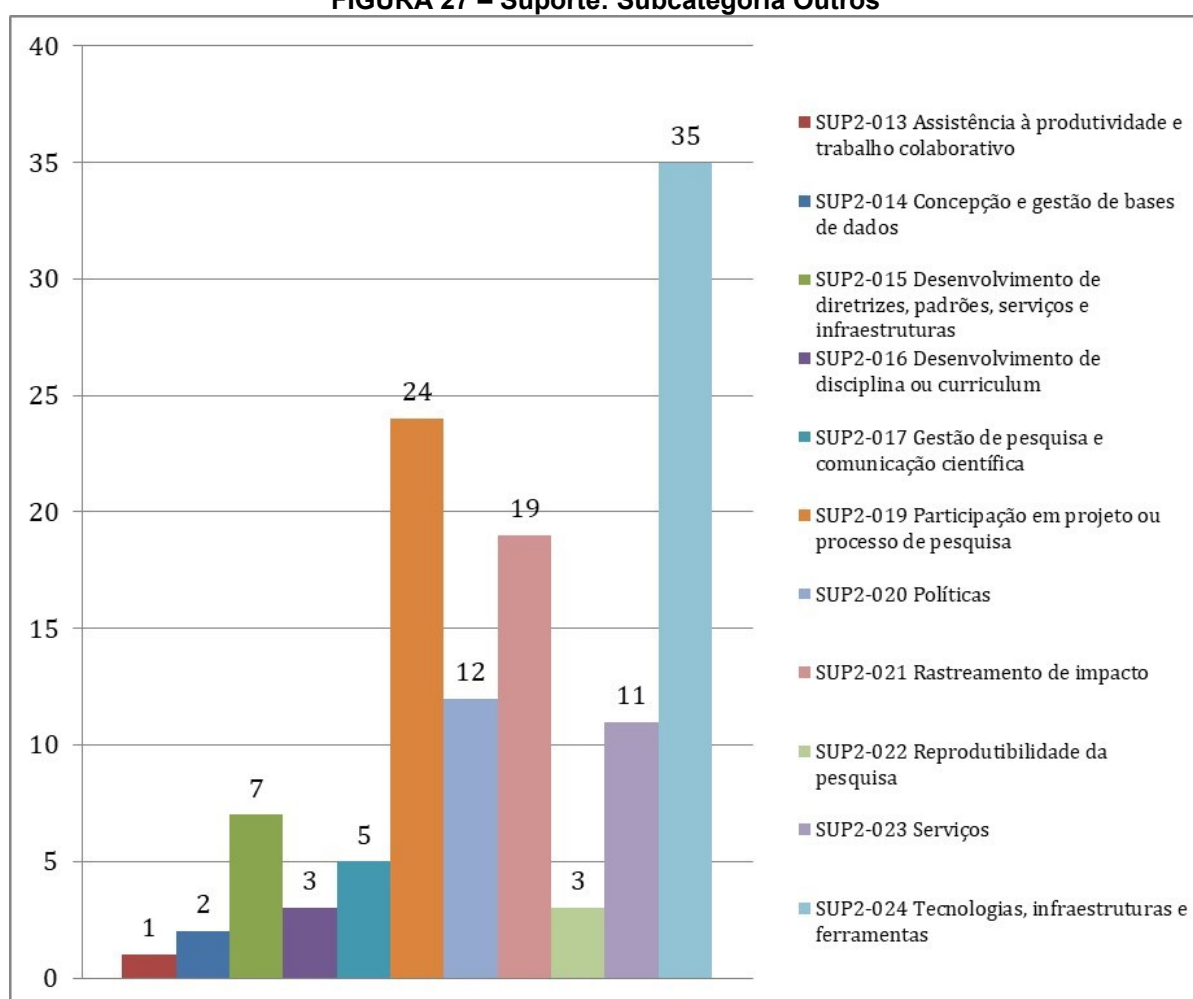
- a) encontrar fontes de financiamento;
- b) pedido de financiamento (escrita e submissão);
- c) planejamento da gestão de dados antes de iniciar a pesquisa, incluindo as estratégias de preservação;
- d) escrita do plano de gestão de dados;
- e) método de pesquisa escolhido;
- f) estruturação da coleta tendo em vista o compartilhamento de dados;
- g) cumprimento dos requisitos para a gestão de dados das agências financiadoras e editores, incluindo requisitos de acesso aberto;
- h) cumprimento dos mandatos políticos.

Na subcategoria ‘Outros’ – SUP2 Outros -, foram encontradas 122 citações aos serviços de suporte, conforme já apresentado na TAB. 2. Suporte a tecnologias, infraestruturas e

ferramentas - SUP2-024 Tecnologias, infraestruturas e ferramentas – destacaram-se em 35 menções (2,17% das referências marcadas na categoria). A subdivisão ‘SUP2-019 Participação em projeto ou processo de pesquisa’ recebeu 24 citações, ou seja, 1,49% das referências marcadas na categoria ‘Suporte’. Descontados os serviços da subdivisão ‘SUP1 000 Gestão e curadoria geral’, os percentuais para ‘SUP2-024 Tecnologias, infraestruturas e ferramentas’ e ‘SUP2-019 Participação em projeto ou processo de pesquisa’ incidiram, respectivamente, 2,67% e 1,83% das ocorrências, conforme já apresentado na TAB. 3.

A FIG. 27 ilustra a quantidade de citações aos outros serviços de suporte ‘SUP2 Outros’.

FIGURA 27 – Suporte: Subcategoria Outros



Fonte: Dados de pesquisa (2019).

Os serviços da subdivisão ‘SUP2-024 Tecnologias, infraestruturas e ferramentas’ compreendem o oferecimento de suporte técnico, assistência, orientação ou consultoria para uso:

- a) Geographic Information System (GIS);
- b) Nvivo;

- c) *software* estatístico;
- d) ferramentas para manipulação de dados;
- e) sistemas de gestão de dados de pesquisa;
- f) *online survey*;
- g) dispositivos móveis;
- h) outras tecnologias, infraestruturas e ferramentas disponíveis.

A subdivisão 'SUP2-019 Participação em projeto ou processo de pesquisa' compreende a participação direta, suporte ou parceria para melhoria das práticas de gestão de dados nos: a) processos; b) projetos; e c) equipes de pesquisa.

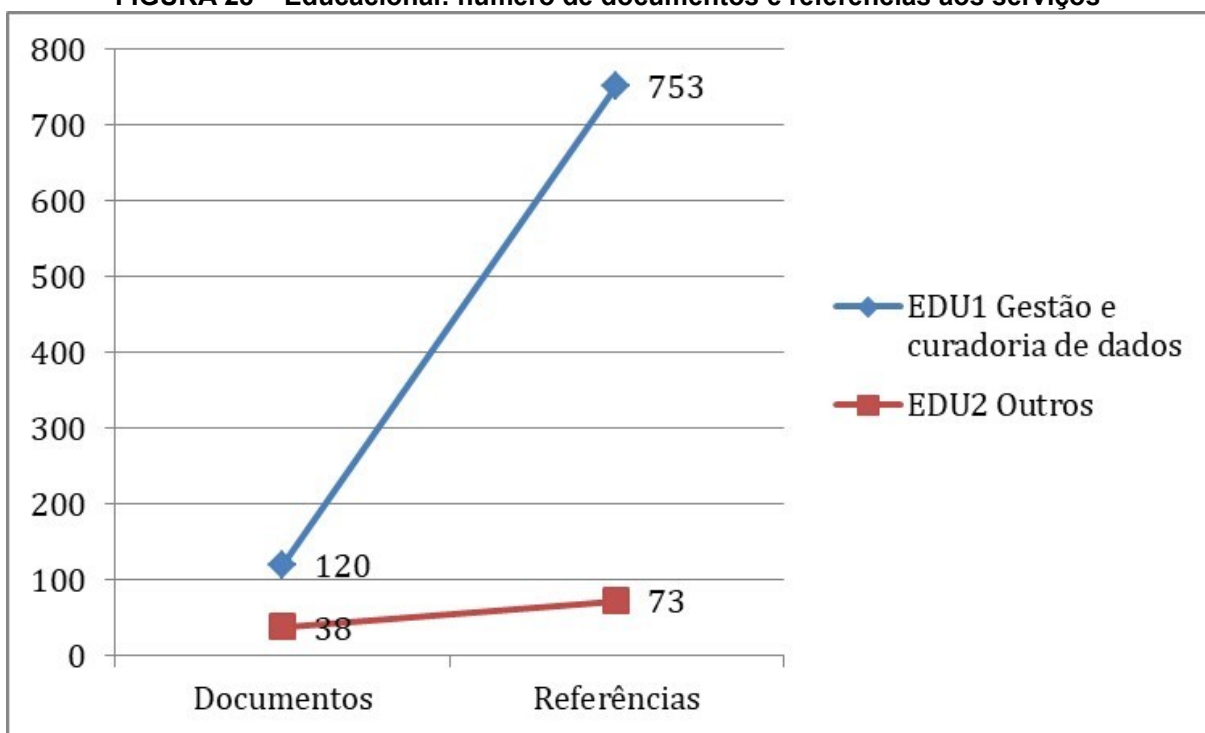
Os serviços da categoria 'Suporte' foram os mais mencionados nos documentos analisados, com destaque para os serviços de suporte ao armazenamento de dados, principalmente relacionados aos repositórios de dados. Merecem destaque também o suporte ao planejamento, sobretudo na escrita e submissão do plano de gestão de dados. Além desses, podemos citar a participação e suporte condução da pesquisa, refletindo uma aproximação da biblioteca com o pesquisador.

A listagem completa dos serviços incluídos na categoria '2 – Suporte' pode ser consultada no Apêndice K.

4.2.3. Categoria Educacional

Os serviços da categoria 'Educacional' destacaram-se nos documentos analisados. Nesta categoria foram considerados serviços como *advocacy*; realização de eventos, tais como: seminários, palestras, simpósios; treinamentos; *workshops*; disponibilização de tutoriais e guias *etc.*

Nesta categoria, 120 documentos fizeram referência aos serviços da subcategoria 'EDU1 Gestão e curadoria de dados' 753 vezes. Outros 38 documentos fizeram alusão aos serviços da subcategoria 'EDU2 Outros' 73 vezes. No total, constatarem-se 826 menções aos serviços desta categoria (FIG. 28).

FIGURA 28 – Educacional: número de documentos e referências aos serviços

Fonte: Dados de pesquisa (2019).

As 826 menções aos serviços da categoria 'Educacional' perfizeram um total de 30,25% das referências aos serviços. Destas 826 referências, 753 ou 91,16% das referências pertencem à subcategoria 'EDU1 Gestão e curadoria de dados' que representa 27,57% do total de serviços nas três categorias. Outras 73 referências, ou seja, 8,84% da categoria ou 2,67% do total de referências marcadas nas três categorias principais (Infraestrutura, Suporte e Educacional), estão presentes na subcategoria 'EDU2 Outros', conforme apresentado na TAB. 4.

TABELA 4 - Educacional: frequência de referências aos serviços

(continua)

3 – EDUCACIONAL		N. de ref. (abs.)	% na categoria	% do total
EDU1 Gestão e curadoria de dados	EDU1-000 Gestão e curadoria – geral	533	64,53	19,52
	EDU1-001 Planejamento	70	8,47	2,56
	EDU1-002 Obtenção	7	0,85	0,26
	EDU1-003 Organização e formato	6	0,73	0,22
	EDU1-004 Documentação de dados e metadados	20	2,42	0,73
	EDU1-006 Armazenamento	45	5,45	1,65
	EDU1-007 <i>Copyright</i> , propriedade intelectual e licenciamento	3	0,36	0,11
	EDU1-008 Questões éticas, legais e de consentimento	10	1,21	0,37
	EDU1-009 Publicação e citação	9	1,09	0,33
	EDU1-010 Uso e reuso	34	4,12	1,24
	EDU1-011 Mineração, análise de dados e visualização	16	1,94	0,59
SUP1 Gestão e curadoria de dados		753	91,16	27,57

TABELA 4 - Educacional: frequência de referências aos serviços

(conclusão)

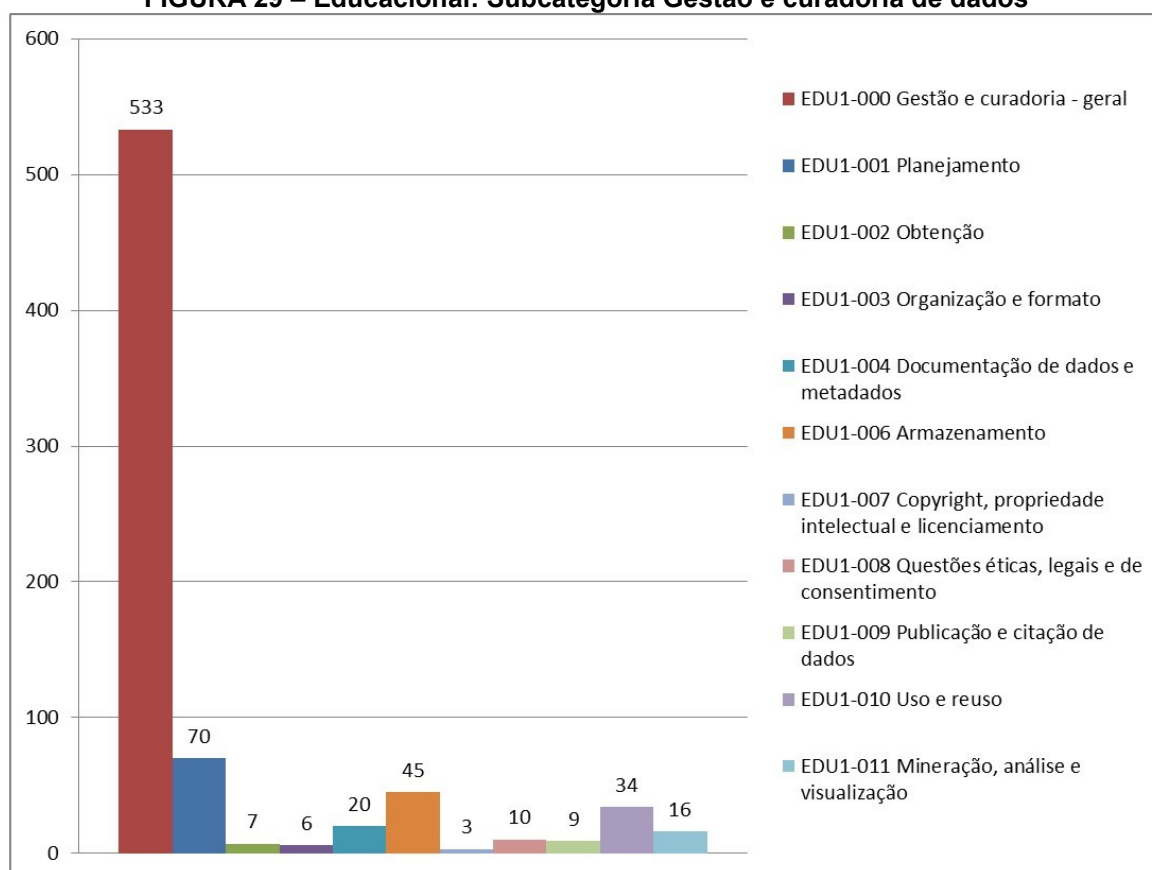
3 – EDUCACIONAL		N. de ref. (abs.)	% na categoria	% do total
EDU2 Outros	EDU2-014 Conceção e gestão de bases de dados	1	0,12	0,04
	EDU2-016 Desenvolvimento de disciplina ou <i>curriculum</i>	10	1,21	0,37
	EDU2-017 Gestão de pesquisa e comunicação científica	1	0,12	0,04
	EDU2-018 Identificação das percepções, práticas e necessidades	14	1,69	0,51
	EDU2-020 Políticas	3	0,36	0,11
	EDU2-022 Reprodutibilidade da pesquisa	2	0,24	0,07
	EDU2-023 Serviços	13	1,57	0,48
	EDU2-024 Tecnologias, infraestruturas e ferramentas	29	3,51	1,06
EDU2 Outros		73	8,84	2,67
Total de referências na categoria 3 - Educacional		826	100,00	30,25
Total geral de referências		2731		

Fonte: Dados de pesquisa (2019).

Assim como na categoria 'Suporte', na categoria 'Educacional, subcategoria 'EDU1 Gestão e curadoria de dados', a maioria dos serviços foi categorizada como gerais – 'EDU1-000 Gestão e curadoria – geral' (FIG. 29), pois esses serviços foram apresentados de forma vaga, imprecisa ou generalizada, conforme trechos a seguir.

[...] offering **workshops and classes on research data management** [...] (GOLDMAN; KAFEL; MARTIN, 2015, p. 5, grifo nosso).

[...] **education services** (CORRALL; KENNAN; AFZAL, 2013, p. 456, grifo nosso).

FIGURA 29 – Educacional: Subcategoria Gestão e curadoria de dados

Fonte: Dados de pesquisa (2019).

Na subdivisão 'EDU1-000 Gestão e curadoria – geral' são encontrados serviços como:

- a) oferecimento de treinamentos, *workshops*, cursos, seminários, conferências, aulas:
 - *data literacy*;
 - gestão de dados de pesquisa;
 - curadoria de dados;
 - melhores práticas de gestão de dados de pesquisa.
- b) desenvolvimento de guias, *templates*, materiais de apoio:
 - gestão de dados de pesquisa;
 - serviços para gestão de dados de pesquisa.
- c) conscientização (*advocacy*):
 - gestão de dados de pesquisa.

Em seguida, a subdivisão mais presente é 'EDU1-001 Planejamento' com 70 menções. Após, aparece a subdivisão 'EDU1-006 Armazenamento'.

Na subdivisão 'EDU1-001 Planejamento' estão presentes serviços como:

- a) oferecimento de treinamentos, *workshops*, cursos, seminários, conferências, aulas:
 - planejamento da gestão de dados;
 - plano de gestão de dados - criação/escrita de planos;
 - uso do DMPTool para criar planos de gestão de dados;
 - requisitos dos financiadores em relação à gestão de dados e ao PGD.
- b) desenvolvimento de guias, modelos, materiais de apoio:
 - plano de gestão de dados - criação/escrita de planos.
- c) conscientização (*advocacy*):
 - plano de gestão de dados.

Na subdivisão 'EDU1-006 Armazenamento' estão presentes serviços como:

- a) oferecimento de treinamentos, *workshops*, cursos, seminários, conferências, aulas:
 - transferência de dados;
 - arquivamento de dados;
 - serviços de arquivamento;
 - armazenamento de dados;
 - opções para armazenamento de dados;
 - compartilhamento de dados;
 - compartilhamento de dados sensíveis;
 - possibilidades e limitações do compartilhamento de dados;
 - práticas e padrões de compartilhamento;
 - *backup* de dados;
 - segurança de dados;
 - repositórios de dados;
 - depósito de dados em um repositório.
- b) desenvolvimento guias, tutoriais, modelos, materiais de apoio:
 - repositórios de dados;
 - serviços de repositório;
 - compartilhamento de dados;
 - requisitos de compartilhamento de dados;

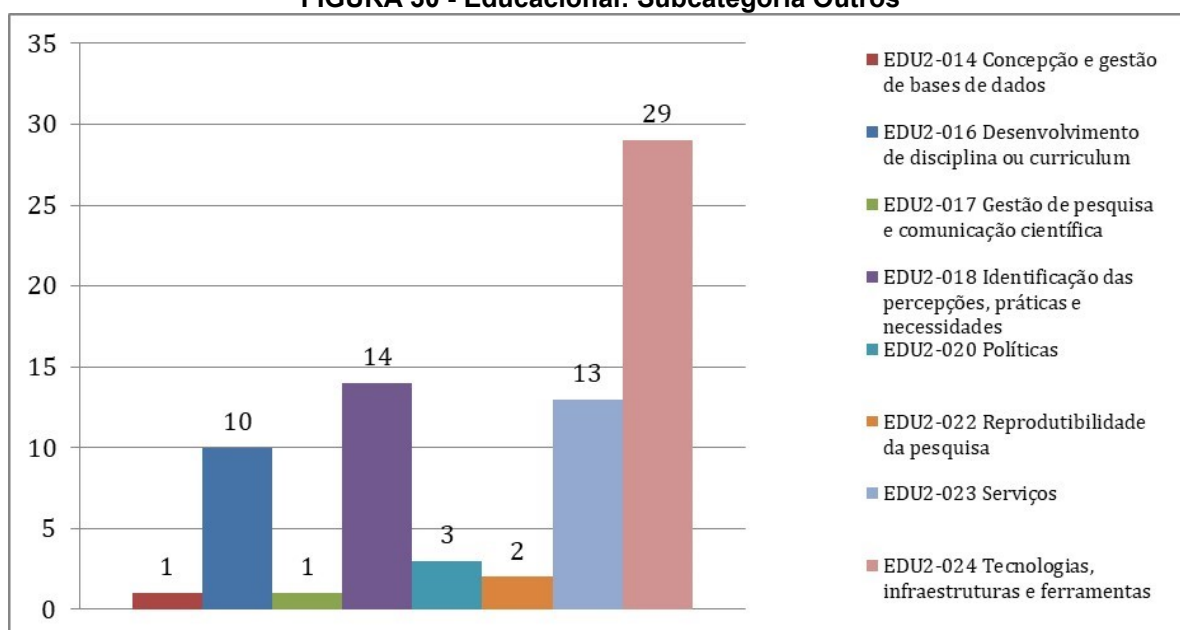
- requisitos de retenção de dados ou conjunto de dados;
- opções de armazenamento de dados.

c) conscientização (*advocacy*):

- requisitos de compartilhamento de dados.

Na subcategoria 'EDU2 Outros', os serviços mais frequentes foram divididos em 'EDU2-024 Tecnologias, ferramentas e infraestruturas', com 29 referências ou 3,51% da categoria 'Educativa', seguidos dos serviços categorizados em 'EDU2-018 Identificação das percepções, práticas e necessidades', com 14 alusões, conforme FIG. 30.

FIGURA 30 - Educativa: Subcategoria Outros



Fonte: Dados de pesquisa (2019).

Na subdivisão 'EDU2-024 Tecnologias, infraestruturas e ferramentas', são encontrados serviços como:

a) oferecimento de treinamentos, *workshops*, cursos, seminários, conferências, aulas:

- ferramentas para análise de dados;
- ferramentas para gestão de dados;
- uso de *software* estatístico;
- análise de redes sociais;
- Nvivo;
- *Geographic Information System* (GIS);
- métodos estatísticos;
- introdução ao R;

- usando o LaTeX;
- SQL;
- programação em Python;
- controle de versão com GIT;
- Bash scripting;
- Unix;
- REDCap (captura de dados);
- uso de uma interface em particular;
- EndNote;
- desenvolvimento de *web application*.

Os serviços da categoria ‘Educativa’ foram amplamente mencionados nos documentos analisados, atrás, somente, das alusões aos serviços da categoria ‘Suporte’.

Nesta categoria destacam-se as ações educativas de cunho generalista realizadas pelas bibliotecas, como o oferecimento de treinamentos e *workshops* sobre a gestão de dados de pesquisa de forma ampla ou, ainda, sobre as melhores práticas para gestão de dados. Contudo, ações mais focadas também são realizadas, principalmente de planejamento, que inclui plano de gestão de dados (criação, ferramentas, desenvolvimento de modelos e conscientização) e ações educativas relacionadas ao armazenamento de dados, como cursos, treinamentos, *workshops* sobre as opções de armazenamento, repositórios de dados *etc*.

A lista completa dos serviços da categoria 3 – Educativa pode ser consultada no Apêndice L.

4.2.4. Subdivisões mais referenciadas

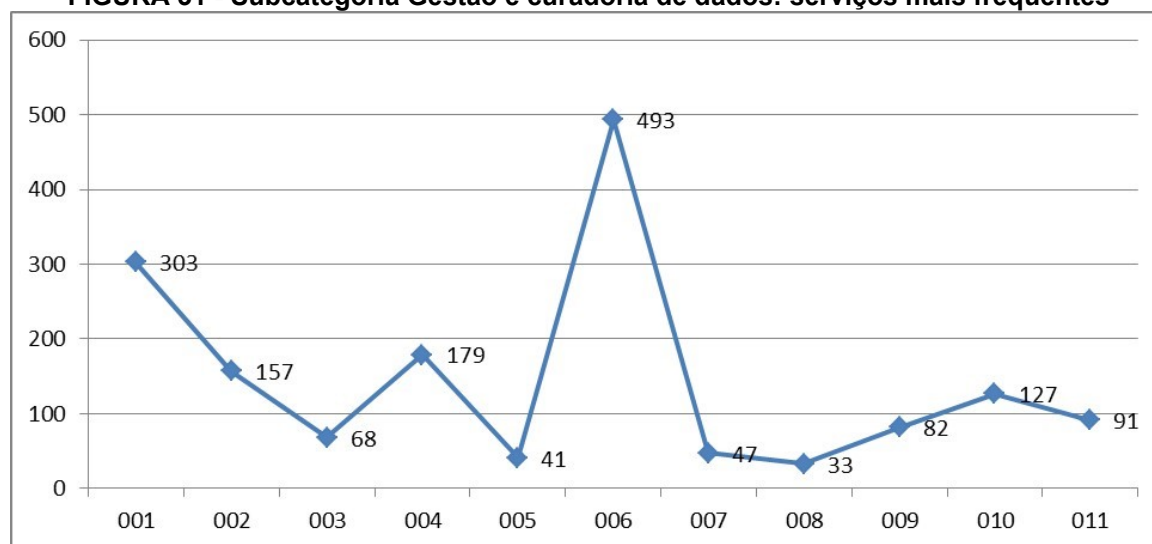
Somando-se as três categorias principais (Infraestrutura, Suporte ou Educativa), observamos que algumas subdivisões se destacaram em termos de menções, conforme TAB. 5.

TABELA 5 - Subcategorias mais referenciadas

SEQ	Nome da categoria	Total
000	Gestão e curadoria – geral	862
006	Armazenamento	493
001	Planejamento	303
004	Documentação de dados e metadados	179
002	Obtenção	157
010	Uso e reuso	127
011	Mineração, análise de dados e visualização	91
009	Publicação e citação	82
003	Organização e formato	68
007	<i>Copyright</i> , propriedade intelectual e licenciamento	47
005	Avaliação e seleção	41
008	Questões éticas, legais e de consentimento	33
Total de referências encontradas		2483

Fonte: Dados de pesquisa (2019).

Pelos motivos já explicados anteriormente, desconsiderando a subdivisão ‘000 Gestão e curadoria – geral’, - serviços não especificados, apresentados de forma generalista, vaga ou ampla –, a subdivisão de mais destaque foi a ‘006 Armazenamento’, seguida por ‘001 Planejamento’ e por ‘004 Documentação de dados e metadados’, conforme FIG. 31.

FIGURA 31 - Subcategoria Gestão e curadoria de dados: serviços mais frequentes

Fonte: Dados de pesquisa (2019).

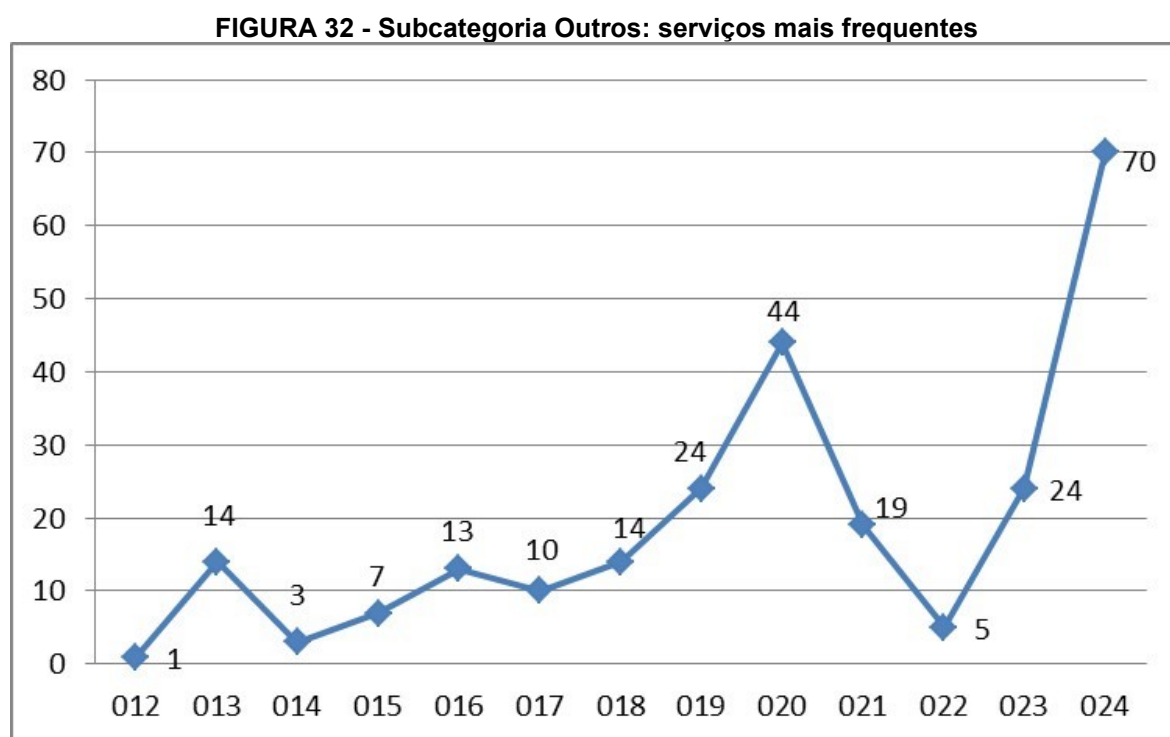
Observamos que os serviços voltados para o armazenamento de dados são os mais requisitados. Eles englobam assuntos como compartilhamento e depósito de dados, repositórios de dados, armazenamento em curto e em longo prazo, arquivamento, serviços de repositórios, identificação de repositórios, *backup* e segurança de dados, criptografia, espelhamento e replicação, eliminação de dados com segurança, integridade dos dados, transferência de dados, dentre outros. Nessa subdivisão, o depósito de dados em

repositórios sobressai, demandando tanto infraestrutura para o armazenamento de dados quanto o suporte técnico e educação para realizar esse depósito.

O planejamento envolve assuntos como: plano de gestão de dados (PGD); planejamento da criação, armazenamento e preservação de dados; questões relacionadas ao financiamento e conformidade com os requisitos dos financiadores e editores. Nessa subdivisão, destacam-se os serviços que apoiam a escrita do plano de gestão de dados. Incluem-se tanto fornecimento de infraestruturas (ferramentas) para elaboração do PGD como os serviços de suporte e educacionais.

A documentação de dados e metadados abrange a descrição de dados, padrões para descrição, atribuição de metadados, geração automática de metadados, dentre outros assuntos. Nessa subdivisão, destacam-se o suporte e serviços educacionais para descrição de dados e atribuição de metadados.

Na subcategoria 'Outros', destacaram-se as subdivisões '024 Tecnologias, infraestruturas e ferramentas' e '020 Políticas', conforme FIG. 32.



Fonte: Dados de pesquisa (2019).

Na subdivisão '024 Tecnologias, infraestruturas e ferramentas', encontram-se serviços relacionados aos outros tipos de tecnologias, infraestruturas e ferramentas oferecidas aos pesquisadores, além daquelas voltadas especificamente à gestão e curadoria de dados, como *survey tools*, redes, dispositivos móveis, LaTeX, dentre outras.

A subdivisão '020 Políticas' abrange as políticas institucionais para a gestão de dados de pesquisa, incluindo questões sobre o compartilhamento dos dados; autoria; acesso; submissão de dados em repositórios; licenciamento para reutilização; período de embargo; normas disciplinares de forma a garantir a integridade da pesquisa; e o cumprimento dos requisitos das agências financiadoras.

4.2.5. Apontamentos gerais

Embora Tenopir *et al.* (2014, 2015, 2017) classifiquem os serviços em informativos/consultivos ou serviços técnicos, na construção da taxonomia, optamos por unir esses dois tipos de serviços na única categoria – Suporte –. A literatura analisada, muitas vezes, não permitia distinguir quando era um serviço técnico realizado pela biblioteca ou um serviço do tipo consultivo, como por exemplo, a identificação de conjunto de dados e a criação de metadados. Esses podem ser serviços técnicos realizados pelas bibliotecas ou serviços de orientação.

Optamos, ainda, diferentemente dos trabalhos de Tenopir *et al.* (2014, 2015, 2017), por incluir a categoria 'Infraestrutura', conforme os tipos de serviços propostos por Reznik-Zellen; Adamick e McGinty (2012), que dividiram os serviços em Educacionais, Consulta ou Infraestrutura. Além disso, a pesquisa de Tenopir *et al.* (2017) descobriu que as bibliotecas acadêmicas europeias fornecem ou participam das infraestruturas tecnológicas oferecendo, principalmente, o armazenamento de dados, seguidos por outros tipos de infraestruturas como as ferramentas para análise de dados e suporte à comunidade virtual.

Os resultados obtidos por Tenopir *et al.* (2017) corroboram este trabalho, uma vez que na categoria 'Infraestrutura', o armazenamento de dados – INF1-006 Armazenamento – é a subdivisão mais mencionada na literatura. Outras infraestruturas também estão presentes, como a subdivisão 'INF1-011 Mineração, análise e visualização' que contempla as ferramentas para análise de dados, e a subdivisão 'INF2-013 Assistência à produtividade e trabalho colaborativo', que abarca as ferramentas de suporte à comunidade virtual.

Dentro da categoria 'Infraestrutura', incluímos as políticas. Qin (2013) indica que as infraestruturas de dados são formadas por tecnologias, padrões e políticas. Outros autores como Rodrigues *et al.* (2010) e Wilson *et al.* (2010, 2011), da mesma forma, indicam as políticas como importantes infraestruturas para a gestão e curadoria de dados.

Tenopir *et al.* (2017) afirmam que um dos serviços mais realizados pelas bibliotecas europeias é o desenvolvimento de políticas institucionais para a gestão de dados.

Na taxonomia, as menções às infraestruturas políticas representam 9,90% das referências aos serviços da categoria. O número é bem representativo porque, excetuando os serviços da subdivisão 'INF1-000 Gestão e curadoria de dados – geral', as políticas – INF2-020 Políticas - só ficam atrás das menções às infraestruturas para armazenamento de dados – INF1-006 Armazenamento -. Tal fato confirma a importância das políticas institucionais para o oferecimento de SSGDPs.

Além das duas categorias citadas anteriormente (Infraestrutura e Suporte), assim como Reznik-Zellen; Adamick e McGinty (2012), optamos por criar a categoria 'Educacional', que abrange serviços para formação em gestão e curadoria de dados como: conferências; seminários; simpósios; palestras e painéis; elaboração e disponibilização de guias; *templates*; tutoriais ou *help*; treinamentos; oficinas; cursos; aulas, dentre outros.

Geraci, Humphrey e Jacobs (2012) defendem que os dados não são diferentes dos outros formatos de conteúdo da biblioteca. Os SSGDPs seriam extensão do papel tradicional da biblioteca, conforme apontado por Tenopir *et al.* (2014, 2015, 2017). A partir da classificação de serviços proposta, percebe-se que os serviços categorizados em 'Suporte' e 'Educacional' são os mais frequentes na literatura, representando, assim, uma extensão dos serviços já realizados pelas bibliotecas. Nestas categorias, estão presentes os serviços que se alinham aos serviços tradicionais de referência e instrução realizados pelas bibliotecas. Essas duas categorias apresentam serviços com grande potencial para aproximação da biblioteca com os pesquisadores, conforme discutido por Monastersky (2013) e ARL (2007).

Os trabalhos de Tenopir *et al.* (2014, 2015, 2017) apontam que os serviços técnicos são os tipos menos oferecidos ou planejados pelas bibliotecas. A menor oferta de serviços técnicos pode ser reflexo dos grandes investimentos que esses serviços requerem, como tempo, recursos e novos conhecimentos técnicos.

Esta pesquisa, porém, descobriu que os serviços de suporte e educacional (extensão dos serviços já realizados pelas bibliotecas) são os mais mencionados na literatura, o que pode sugerir que a sua implementação requer menos investimentos e, portanto, seriam mais fáceis de serem realizados pelas bibliotecas.

4.3. CONTEXTUALIZAÇÃO NAS INSTITUIÇÕES ENTREVISTADAS

Esta seção apresenta a contextualização dos serviços e infraestrutura para gestão de dados nas duas instituições entrevistadas: a Universidade Federal de Santa Catarina (UFSC) e a Universidade Federal do Rio Grande do Sul (UFRGS).

4.3.1. Gestão de dados de pesquisa na UFSC

A UFSC ainda não conta com serviço de suporte à gestão de dados de pesquisa ou infraestrutura para depósito de dados estruturado na instituição. Contudo, há serviços em fase de desenvolvimento.

[...] a gente ainda não têm o serviço propriamente, ou a gama de serviços, porque a nossa ideia é que seria uma gama de serviços (P1, 2019).

A discussão, porém, sobre a gestão de dados na UFSC, começou a partir de um estágio que a entrevistada, bibliotecária da UFSC, fez na Universidade Tecnológica de Munique na Alemanha no ano de 2015.

Mas a origem, a gente começou a discutir isso aqui foi em 2015. Na ocasião eu fiz uma viagem pra Alemanha, foi uma espécie de estágio. Lá, um dos serviços, umas das bibliotecas que eu fiquei mais em contato, eles já tinham criado um suporte à pesquisa e gerenciamento de dados, um serviço relacionado. Eles não gostavam muito de ficar enfatizando a questão do gerenciamento de dados. Eles tinham essa visão de que é algo mais amplo e que é um serviço de suporte à e-pesquisa, pesquisa eletrônica ou usando os recursos eletrônicos (P1, 2019).

A partir da experiência adquirida na Alemanha, foi criada uma comissão para desenvolvimento dos serviços de suporte à pesquisa, posteriormente formalizada por meio de portaria da Universidade, no ano de 2016.

Quando eu voltei dessa viagem, [...] uma questão que estava envolvida com essa minha oportunidade de ir para lá era trazer alguma novidade para a minha instituição. Então a gente buscou se reunir e até que surgiu a comissão para concepção dos serviços de suporte à pesquisa (P1, 2019).

[...] se for pensar que a comissão existe desde 2015, basicamente 2016, porque 2015 eu voltei de viagem. Já era segundo semestre. Acho que até a formalização mesmo pela portaria era 2016 (P1, 2019).

A comissão foi formada, além da biblioteca, por outras instâncias como a Pró-Reitoria de Pesquisa (PROPESQ), Pró-Reitoria de Pós-Graduação (PROPG), Superintendência de Governança Eletrônica e Tecnologia da Informação e Comunicação (SeTIC) e Secretaria de Inovação (SINOVA).

[...] a gente já tinha consciência assim, da necessidade do envolvimento das outras instâncias. Então, foi convidada a pró-reitoria de pesquisa, a pró-reitoria de pós-graduação, a superintendência de tecnologia da informação da universidade, o escritório de inovação (P1, 2019).

A comissão reúne-se periodicamente e, como uma de suas primeiras ações, realizar levantamento dos comportamentos e necessidades dos pesquisadores da UFSC. Assim, foram escolhidos, como amostra da pesquisa, os pesquisadores bolsistas de produtividade do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

[...] umas das primeiras ações da nossa comissão foi fazer esse diagnóstico interno sobre o que os nossos pesquisadores estão fazendo, qual é o comportamento deles, o que eles esperam de um serviço de suporte à pesquisa e gestão de dados. E, aí, a gente fez um questionário internamente, com a pesquisa mesmo, foi até colocado também para o Comitê de Ética [...] (P1, 2019).

[...] nessa pesquisa a gente conseguiu ter um levantamento, não foi tão [...] abrangente como a gente gostaria, porque [...] gente resolveu fazer só com os pesquisadores que eram bolsistas de produtividade [...] (P1, 2019).

A entrevistada destacou a importância de encontrar evidências sobre: saber como os pesquisadores fazem a gestão dos seus dados, como acontece na prática, assim, “[...] por mais que a gente soubesse assim né, de conversas, como que é, a gente queria ter, documentado [...]” (P1, 2019).

[...] a gente gostaria de ter tido mais respostas, mas já foi bem esclarecedor. A gente tinha evidências, agora, para afirmar para a instituição: olha, nosso pesquisador precisa de infraestrutura tecnológica, precisa de espaço em disco, precisa de uma política para definir se ele é o dono daquela informação ou se é a instituição (P1, 2019).

Além das práticas e necessidades dos pesquisadores, o levantamento permitiu identificar as lacunas a serem resolvidas antes de tomar qualquer decisão sobre a criação de uma infraestrutura para o depósito de dados. Foi identificado que a instituição precisa de uma política informacional de acesso aberto. Precisa, também, definir quem é o dono dos dados, a organização ou o pesquisador. Esses pontos permanecem confusos para os pesquisadores.

Precisa de política informacional de acesso aberto, sabe? Então, a gente conseguiu, com esses dados, perceber [...] lacunas que a gente precisaria preencher antes de qualquer decisão de [...] abrir um repositório de dados [...] Porque a gente não quer fazer isso! Vamos abrir e pronto! Agora a gente tem um repositório de dados. A gente vai recebendo os dados e vai pensando no que fazer. Não era essa a ideia (P1, 2019).

[...] uma política para definir se ele [o pesquisador] é o dono daquela informação ou se é a instituição. Isso tem que estar claro pra ele [pesquisador]. Para muitos pesquisadores não está claro (P1, 2019).

Uma segunda ação da comissão foi a realização de um evento, que objetivou discutir o assunto e servir como prestação de contas, mostrar aos pesquisadores o que a comissão havia realizado. A entrevistada afirma que o “[...] objetivo era mostrar para os nossos pesquisadores o que a gente tinha feito até então” (P1, 2019).

[...] em virtude dessa comissão, [...] a gente organizou aquele seminário, [I Seminário de Suporte à Pesquisa e Gestão de Dados Científicos que ocorreu nos dias 18 e 19 de setembro de 2017], há dois anos [...] (P1, 2019).

A entrevistada afirma que, de palpável, começaram essas duas ações: o levantamento do comportamento e necessidade dos pesquisadores e o I Seminário de Suporte à Pesquisa e

Gestão de Dados Científicos. A entrevistada considera que os trabalhos da comissão avançaram pouco, uma vez que eles se reúnem desde 2016, quando a ideia foi apresentada à gestão da instituição. Contudo, apareceram muitos percalços “[...] se for pensar nisso, que está [a comissão] desde 2016, a gente avançou muito pouco, mas porque teve muito percalço” (P1, 2019).

[...] a gente conseguiu apresentar a ideia. De início a gente teve [...] uma abertura boa, só que, também, [...] teve muito percalço institucional [...] a UFSC passou por uns momentos bem complicados nesses últimos anos, troca de gestão [...] inesperadas [...] (P1, 2019).

Além dos percalços que dificultaram o avanço das ações, principalmente com a troca inesperada do reitorado da instituição, a UFSC decidiu realizar todo o processo de forma gradual. A entrevistada afirma que a comissão “[...] tinha a intenção de fazer as coisas gradativamente e com suporte certo, sem achismo e sem [...] se aventurar, sem saber direito onde estava pisando”. Além disso, a entrevistada disse que sempre pensou “[...] na necessidade de primeiro ter que estabelecer as políticas para, depois, [executar] as próximas etapas”, e após isso, pensar em suporte tecnológico (P1, 2019).

Como é que a gente vai falar para o pesquisador: olha, tu podes depositar os dados aqui no nosso repositório, mas ele vai questionar tudo isso. É aberto ou não é aberto? O que eu posso ter de privacidade? O que eu não posso? (P1, 2019).

Atualmente a comissão está trabalhando numa minuta de política informacional de acesso aberto para a instituição.

O que a gente está trabalhando e que [esperamos] terminar na próxima reunião [...] é que, [...] por mais que a gente pense na questão tecnológica, nos serviços em si, não dá pra pensar em muita coisa sem definir políticas [...] (P1, 2019).

[...] a nossa instituição não tem políticas claras, nem para suas publicações, muito menos para dados [...] [Estamos] trabalhando numa minuta de política de acesso aberto para a instituição, [...] Por enquanto, a comissão está fazendo essa minuta e, assim que ela terminar, [será] levada à discussão nas instâncias que precisam, [...] nas pró-reitorias, no conselho universitário, enfim, vai ter, claro, outras discussões que vão sair do âmbito até da própria comissão [...] (P1, 2019).

As próximas etapas de trabalho da comissão abordarão a discussão da infraestrutura tecnológica, as parcerias com outras instituições (Universidade Tecnológica de Munique, RNP, IBICT), a capacitação e formação das equipes para o suporte e a definição das funções das bibliotecas e das outras instâncias envolvidas.

Esclarecendo melhor [...] **Terminou a política, vem a tecnologia, às vezes, quase que em conjunto vem a questão da definição de habilidades, competências, serviços** (P1, 2019, grifo nosso).

[...] as **próximas etapas**, eu imagino, [seriam] **começar a discutir a questão tecnológica**, [...] **a infraestrutura, qual repositório [...] usar**. [...] A gente usa o DSpace. **Ou vamos ter um [repositório] separado e outro software** [...] para dados? **Vamos fazer parceria com [a] instituição da Alemanha**, por exemplo? [...] A gente sabe que a **RNP** [Rede Nacional de Ensino e Pesquisa] também está trabalhando nessa questão. Outras instituições, como o **IBICT** [Instituto Brasileiro de Informação, Ciência e Tecnologia também está trabalhando.] [...] **Depois, acho que é muito importante, a capacitação das equipes, porque, quem que vai dar esse suporte?** (P1, 2019, grifo nosso).

O que será função da biblioteca? O será função de outro órgão? [...] Por exemplo, a pró-reitoria de pesquisa já tem alguns serviços específicos. Então, **tem coisas que, devemos, pelo menos, saber direcionar**. Isso aqui é para a Propesc [Pró-reitoria de Pesquisa]. Isso aqui é para [a biblioteca]. Isso aqui é para outra instância. Isso é relacionado a direito autoral, vai para a Secretaria de Inovação [...] Então, **tem toda uma formação da equipe da biblioteca em consonância com as outras instâncias** para poder fazer [o trabalho] (P1, 2019, grifo nosso).

De acordo com a entrevistada, terminadas as definições políticas, surgem outras questões e, depois, a comissão pretende criar um projeto piloto.

[...] **terminou a política, vem a tecnologia**, às vezes, quase que em conjunto vem a **questão da definição de habilidades, competências, serviços**, então, assim, tem muitas coisas para ser ainda, discutidas na comissão [...] **a ideia, inclusive, é fazer um projeto piloto com um, dois grupos de pesquisa** [...] (P1, 2019, grifo nosso).

A entrevistada lembra que, a partir do momento em que a instituição conseguiu “[...] conceber soluções para a gestão de dados e suporte à pesquisa mais próximo ao nosso pesquisador”, surgem outras necessidades, como a manutenção dos serviços (P1, 2019).

A entrevistada vê os serviços de suporte à gestão de dados ou suporte à pesquisa como continuidade dos serviços que a biblioteca já oferece, com o objetivo de se manter relevante perante a comunidade de usuários.

[...] essa foi a minha visão. Quando eu tive contato, talvez **por causa da gestão dos dados brutos, parece que [o suporte à gestão de dados] é uma novidade** [...] **A minha percepção é de continuidade. É uma ampliação** [...] (P1, 2019, grifo nosso).

Sabe, eles têm tudo! Então **eles** [Universidade Tecnológica de Munique] **questionavam: [...] como que a gente ainda pode se reinventar, se colocar perante nosso usuário, nossa comunidade e permanecer relevante? É realmente uma continuidade do que a gente já faz, de uma maneira que [...] possamos permanecer relevantes**. Às vezes, o pesquisador não imaginaria que a biblioteca poderia atuar naquela questão, por exemplo. Mas **eu vejo como continuidade daquilo que classicamente, uma biblioteca universitária ou de pesquisa já faz** (P1, 2019, grifo nosso).

4.3.2. Gestão de dados de pesquisa na UFRGS

A UFRGS participa de dois projetos relativos à gestão de dados de pesquisa: em âmbito nacional (1), onde a Universidade participa do Grupo de Trabalho da Rede de Dados de Pesquisa Brasileira (RDP Brasil), em colaboração entre a Rede Nacional de Ensino e Pesquisa (RNP) e o Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT), além da Universidade do Rio Grande (FURG). O projeto tem como objetivo a criação de um Programa Nacional de Acesso Aberto a Dados de Pesquisa; em nível local (2), projeto que trata do desenvolvimento de um repositório de dados para a Universidade.

O projeto da RNP e IBICT é desenvolvido pelo Centro de Documentação e Acervo Digital da Pesquisa (CEDAP) da Faculdade de Biblioteconomia e Comunicação. O repositório de dados da UFRGS é um projeto do Centro de Processamento de Dados (CPD).

[...] esse [o repositório de dados da UFRGS] é um projeto lá do CPD, do Centro de Processamento de Dados da UFRGS, o outro [da RNP e IBICT], é o projeto aqui da Faculdade de Biblioteconomia, do CEDAP, [Centro de Documentação e Acervo Digital da Pesquisa, Órgão Auxiliar da Faculdade de Biblioteconomia e Comunicação] que é esse projeto da RNP e do IBICT, que tem um âmbito nacional (P2, 2019).

A entrevistada conta que atuou como bibliotecária no CPD por 25 anos e atualmente é docente da Faculdade de Biblioteconomia e Comunicação. Assim, compreende-se o motivo de os projetos terem surgido a partir dessas duas instâncias na Universidade. Ela aponta que atua nos dois projetos “[...] eu trabalho nos dois. Em um, que é específico da Universidade, e outro, que é esse em âmbito nacional, da RNP e do IBICT” (P2, 2019). A participação simultânea nesses dois projetos “[...] facilita as coisas, porque tudo que eu faço aqui [UFRGS] serve para lá [Projeto RNP e IBICT], e tudo que eu faço lá, serve para cá, então facilita bastante” (P2, 2019).

Segundo a entrevistada, o projeto da RNP (primeiro projeto) possui três objetivos:

1. Identificação dos repositórios de dados nacionais, realizado por meio de levantamento no re3data³⁵;

A gente, na primeira etapa, [...] tinha alguns objetivos para alcançar. Esses objetivos eram: **fazer um levantamento dos repositórios nacionais**, que repositórios eram esses, o que eles continham, de que áreas eles eram, qual o *software* utilizavam, quais as instituições estavam por trás desses repositórios. **[Fizemos] um levantamento pelo re3data** e começamos a estudar esses repositórios (P2, 2019, grifo nosso).

³⁵ Registry of Research Data Repositories, um registro global para repositório de dados de pesquisa.

2. Identificação das práticas dos pesquisadores, por meio do envio de 71 mil *e-mails* para pesquisadores e para 2400 coordenadores de pós-graduação, sendo recebidas 4735 respostas;

O segundo objetivo era fazer um levantamento nacional para identificar as práticas com os repositórios de dados científicos ou as práticas dos pesquisadores em utilizar repositórios e também, [a utilização de] [...] dados de outros repositórios, entre outras coisas (P2, 2019, grifo nosso).

Descobrimos várias coisas: como os pesquisadores colaboram uns com os outros; onde eles depositam [seus dados]; quem eles [acreditam] que deve ser o responsável pela infraestrutura para o depósito de dados científicos (P2, 2019).

3. Avaliação das ferramentas para repositório de dados, identificadas a partir dos repositórios encontrados.

[...] a partir dos repositórios que já tinham, nós identificamos que usavam Dataverse, DSpace, CKAN e depois nós fizemos a avaliação do Zenodo. Então foram instalados aqui no laboratório todos para aprender a usar e entender cada um deles e ver as características de cada um deles (P2, 2019).

A gente avaliou diversas características, acho que umas 20 características, para ver qual deles tinha todas as características que a gente considerava essenciais para uma ferramenta para repositório de dados científicos. O CKAN foi o que menos teve os princípios cobertos e o Dataverse foi o que mais teve. O DSpace ficou faltando uns 3 ou 4 critérios [princípios essenciais para uma ferramenta para repositório de dados de pesquisa] (P2, 2019).

O segundo projeto é a criação de um repositório de dados na UFRGS. Está sendo desenvolvido a partir das informações coletadas no projeto de âmbito nacional da RNP e IBICT. Assim, a UFRGS está em fase de desenvolvimento de um protótipo do repositório de dados usando o Dataverse, ferramenta escolhida por apresentar as características essenciais definidas no âmbito do projeto.

Um protótipo onde a gente optou pelo Dataverse para criar repositórios de dados. Tem três Dataverses, de três dados [tipos de dados]. Escolhemos dados de redes de computadores e mais os dados nossos [daqui] do CEDAP, [os] dados do nosso levantamento dos 70 mil *e-mails*. Quero dizer, os 4 mil e tantas [respostas] (P2, 2019).

Após isso, pretende-se criar um agregador de repositórios ou comunidade federada.

E agora, o projeto está numa fase 2, que é outro trabalho, usando o que já fizemos. Entrar numa fase 2, que seria de um agregador de repositórios ou uma comunidade federada, isso ainda não está certo (P2, 2019).

Podemos considerar que o desenvolvimento do repositório de dados teve como antecedente o repositório institucional Lume. A entrevistada trabalhou [...] com o repositório institucional e

depois [...], há uns 2 anos [...] fez um projeto para desenvolvimento do repositório de dados (P2, 2019).

[...] na Universidade, eu sempre trabalhei com banco de dados, com a gestão do banco de dados das 32 bibliotecas. Nos últimos 10 ou 11 anos, [trabalhei] com o repositório [institucional Lume]. Na verdade, [...] uns 15 anos, porque [comecei] a estudar bem antes (P2, 2019).

Tanto o repositório institucional Lume quanto o projeto do repositório de dados não surgiram das bibliotecas, mas de iniciativa de bibliotecários que atuavam no Centro de Processamento de Dados da UFRGS.

E, o repositório surgiu como uma iniciativa lá dentro do Centro de Processamento de Dados. Não foi uma iniciativa [...] que saiu das bibliotecas [a partir de] uma necessidade. A gente fez com que ele se fizesse necessário (P2, 2019).

Na UFRGS, a decisão de criação do repositório de dados não foi imposição externa ou de órgãos financiadores, surgiu da necessidade percebida pelos bibliotecários que atuavam no CPD.

[...] foram iniciativas no Centro de Processamento de Dados. Da mesma forma, o projeto do repositório de dados científicos. Não surgiu das bibliotecas ou foi uma solicitação para o Centro de Processamento de Dados. Sempre é o contrário, talvez porque tenha bibliotecários lá! Claro! Porque senão, eles não iriam atinar [para] essas coisas. Então, como sempre, teve bibliotecário ali trabalhando [...] gente sempre atenta com o que está acontecendo [...] (P2, 2019).

Embora não seja iniciativa das bibliotecas, a entrevistada destaca que o trabalho só foi possível devido às atividades desenvolvidas pelas bibliotecas, uma vez que o “[...] repositório [Lume] não é por autoarquivo”. Toda a produção científica é registrada na biblioteca no Sistema Integrado de Bibliotecas Aleph e, após isso, o repositório Lume (DSpace) é automaticamente alimentado (P2, 2019).

Para o desenvolvimento do repositório de dados, a UFRGS aproveitou as informações obtidas por meio do levantamento das necessidades dos pesquisadores realizado pelo projeto da RNP. Esse levantamento, contudo, teve caráter nacional. Foram filtradas apenas as respostas dos pesquisadores ligados à UFRGS.

Tinham muitas respostas de pesquisadores da Universidade. Filtramos isso e eu levei para o CPD. Dali, selecionamos alguns pesquisadores para [...] entrevistas mais personalizadas. Queríamos ver exatamente que tipo de necessidade eles tinham, quais os tipos de dados [produziam]. Estamos conversando agora diretamente com os pesquisadores [selecionados] [...] desses 4700 [4735 respostas obtidas no levantamento nacional] (P2, 2019).

O Centro de Processamento de Dados, unidade responsável pelo desenvolvimento do repositório de dados, conta com a atuação conjunta de bibliotecários, analistas e

programadores, representando uma junção de competências para a entrega do resultado final. Conforme apontado, “a gente tem analistas de sistemas e técnicos programadores. Um grupo de analistas, programadores e bibliotecários” (P2, 2019).

A UFRGS ainda não definiu as políticas que regem a gestão de dados de pesquisa na Universidade, nem a política para o repositório de dados.

Assim que a gente tiver delineado o repositório, feito testes, teremos que pensar na política. Vamos ter que chamar bibliotecários de diversas áreas, pesquisadores de diversas áreas, para recomençar o processo, assim como fizemos com o repositório institucional. Envolver as pessoas para que essas pessoas repliquem [os fatos] e o repositório funcione (P2, 2019).

A entrevistada destaca que, com o repositório Lume, foi necessário o envolvimento de todos os bibliotecários, desde a elaboração das políticas, direitos autorais, treinamentos, além do funcionamento do próprio repositório, pois, “[...] 80% do que tem no Lume é trabalho dos bibliotecários nas bibliotecas setoriais”, que registram a produção científica na biblioteca e que, por sua vez, é alimentada automaticamente no DSpace a partir de sistema de gerenciamento de bibliotecas Aleph (P2, 2019).

[...] tivemos que envolver todos os bibliotecários, porque algumas coisas [precisaram] ser adaptadas para atender as necessidades dos metadados no DSpace.

No quesito dificuldades encontradas para o desenvolvimento de infraestruturas ou serviços para a gestão de dados de pesquisa, a entrevistada apontou dificuldades trazidas pelo fato de ser área de atuação nova. A equipe não dispunha de conhecimento prévio sobre o trabalho a ser realizado. As soluções ou respostas às indagações eram procuradas na literatura. Algumas questões precisaram, ainda, ser entendidas como: formatos de dados; ferramentas que devem ser incorporadas para garantir a acessibilidade aos dados; a importância do plano de gestão de dados; e a organização dos dados, para que eles possam ser depositados nos repositórios.

A dificuldade é [...] conhecer esse mundo totalmente diferente [...] Por exemplo, o levantamento desses 70 e tantos mil *e-mails*, **a gente levantou vários trabalhos internacionais** [identificação das necessidades dos pesquisadores que faziam isso] também para dados de pesquisa (P2, 2019, grifo nosso).

Na verdade a maior dificuldade, no início, fomos nós mesmos entendermos o que era esse mundo. Encaramos questionários de [outros] trabalhos [...] [e tentamos] entender o que eles queriam perguntar, quando, eu falei: [...] o que são formatos? **O que são formatos de dados?** A gente entende de formato Word, PDF e tal, mas tem milhares de formatos. Tem milhares de **ferramentas que [podem] ser incorporadas aos dados para que [...] possam ser acessíveis.** Não adianta ter dados, [...] que foram coletados no sistema x e ninguém conseguir olhar porque tem que ter um *software* caríssimo para analisar (P2, 2019, grifo nosso).

[...] **tem várias coisas que a gente teve que aprender [...], a questão do plano de gestão de dados** e a importância disso desde, o começo da pesquisa. Acho que esse vai ser o maior problema com os pesquisadores, deles organizarem, desde o plano de gestão de dados até o dado chegar no repositório. **Os dados têm que ser organizados** (P2, 2019).

A entrevistada entende que o maior problema está no fato de os pesquisadores precisarem organizar os dados, desde o plano de gestão até o depósito em repositórios.

Em se tratando do apoio institucional, na UFRGS o CPD, órgão responsável pelo desenvolvimento do repositório de dados, é ligado à Reitoria, o que facilita os trabalhos desenvolvidos e garante apoio institucional, como já aconteceu com o repositório Lume.

[...] o Centro de Processamento de Dados é ligado diretamente à Reitoria. Às questões políticas são mais fáceis de serem resolvidas. No repositório institucional, a gente levava todas as necessidades para a direção do Centro, e a direção encaminhava direto para o reitor. Isso facilitou muito (P2, 2019).

Em outra questão apontada, a entrevistada acredita que alguns pesquisadores nunca irão colocar os dados no repositório da instituição, pois, podem, já ter onde depositar os dados, ou ainda porque uma determinada revista ou financiadora obrigue o pesquisador a depositar os dados em repositório dela.

[...] a gente descobriu, entrevistando alguns pesquisadores, coisas que a gente já via pela literatura. Agora [temos] certeza. Tem pesquisadores que nunca vão colocar os dados no nosso repositório. O dado trabalhado, aquele que foi, sei lá, publicado na Nature³⁶. Os dados estão em repositório fora do Brasil. Tem pesquisadores, por exemplo, da área de genoma, biologia molecular, essas áreas assim *tops*. [Elas] já têm repositórios específicos [para os] dados. E eles [os pesquisadores] não vão trazer os dados para cá, porque a revista exige que ele deposite em um repositório X, por exemplo, nos Estados Unidos (P2, 2019).

A entrevistada destacou que esses pesquisadores, embora depositem os dados finais em algum outro repositório específico, precisarão depositar dados intermediários ou brutos em algum lugar, em uma infraestrutura intermediária que seria o repositório de dados da instituição.

Mas o que ele precisa? [...] De um repositório para colocar os dados intermediários, os dados brutos, aqueles dados com que ele vai trabalhar, para depois, gerar aquele pedacinho só de dados para seu artigo. Isso também é uma coisa que a gente descobriu na literatura e realmente é assim que acontece. A gente entrevistou um professor da biologia molecular. Ele deu uma aula de repositórios de dados científicos, porque essa é uma área do conhecimento que já está acostumada com isso. Tem áreas que vão precisar de infraestrutura intermediária e não o dado final [...] Em outras áreas eles não têm nada. Vão precisar desde [colocar no repositório] o dado bruto até o dado final, o que vai para o artigo [...] (P2, 2019).

³⁶ www.nature.com

4.3.3. Síntese e reflexões: entrevistas

As duas instituições, UFSC e UFRGS, apresentam estágios iniciais de disponibilização de SSGDPs. Optaram por caminhos distintos para proposição inicial de serviços. A UFSC dedicou-se a formação de uma comissão para estudar o assunto. Essa comissão realizou levantamento das necessidades dos seus pesquisadores e um evento de conscientização sobre a necessidade da gestão adequada dos dados de pesquisa e os benefícios trazidos aos pesquisadores e instituição. No momento da entrevista (abril de 2019), a comissão trabalha no desenvolvimento de uma política para a instituição. Depois disso, conforme a entrevistada, irão se dedicar às infraestruturas tecnológicas e a capacitação e formação da equipe.

A UFRGS iniciou o diálogo sobre a gestão de dados de pesquisa da instituição juntamente com outras instituições, por meio da colaboração no projeto desenvolvido pela RNP e IBICT. Dentre os objetivos do projeto, foi realizado levantamento das práticas e necessidades dos pesquisadores em nível nacional e estudo de várias ferramentas para repositório de dados. A UFRGS lançou recentemente um protótipo de repositório para dados de pesquisa.

Consideramos que, diferentemente da UFSC, a UFRGS está trabalhando na definição das infraestruturas tecnológicas como ponto inicial para a disponibilização dos SSGDPs. Contudo, as duas instituições realizaram o levantamento das práticas e necessidades dos pesquisadores. Essa estratégia corrobora o pensamento de diversos autores como Auckland (2012), Gabridge (2009), Jones, Pryor e Whyte (2013) e Tenopir *et al.* (2015), que afirmam que o levantamento das práticas e necessidades busca entender às necessidades dos pesquisadores e desenvolver estratégias específicas para atender à demanda desse público. Coates (2014) afirma que esta etapa é crucial para o desenvolvimento de serviços relevantes.

Corroborando a literatura, constatou-se o fato de a responsabilidade pelo planejamento e desenvolvimento de serviços variar de biblioteca para biblioteca, como apresentado por Tenopir *et al.* (2015). Pudemos observar que, na UFSC, os serviços começaram a ser pensados a partir de uma espécie de estágio em uma instituição alemã realizado por uma bibliotecária do quadro de servidores da Biblioteca Universitária. A partir daí, foi criada uma comissão para discutir os assuntos e implementar as ações. Na UFRGS, a discussão sobre a gestão de dados de pesquisa iniciou a partir da percepção de uma bibliotecária (atualmente docente) da instituição, sem atuação diretamente dentro da biblioteca.

As duas entrevistas permitiram a observação de pontos importantes para o desenvolvimento dos serviços e infraestruturas, conforme fundamentação teórica apresentada neste trabalho:

- a) discussões sobre a gestão de dados a partir da percepção de mudanças no cenário informacional;

As duas instituições relatam que as iniciativas não surgiram de uma obrigação ou recomendação das agências de financiamento, mas da percepção do cenário mundial e da oportunidade de as bibliotecas se manterem relevantes, oferecendo serviços que vão de encontro às necessidades dos pesquisadores.

- b) comissão ou projeto formal para discussão da gestão de dados;

Na UFSC, o desenvolvimento de serviços e infraestrutura é realizado pela biblioteca, mas houve a criação de uma comissão com participação de diversas outras instâncias diretamente ligadas. Na UFRGS, os dois projetos (repositório de dados e o projeto da RNP) são desenvolvidos pelo Centro de Processamento de Dados e pela Faculdade de Biblioteconomia e Comunicação, respectivamente, garantindo a formalização necessária das ações.

- c) levantamento das práticas e necessidades dos pesquisadores;

Tanto UFSC quanto UFRGS realizaram o levantamento das práticas e necessidades dos pesquisadores. Esse tipo de levantamento é uma das primeiras e mais importantes etapas para o oferecimento de serviços bibliotecários relevantes à comunidade.

- d) apoio da alta administração;

As duas instituições reconhecem a importância do apoio da alta administração. UFSC informou que obteve apoio e abertura para o projeto, mas a instituição passou por mudanças inesperadas no reitorado, o que pode mudar as prioridades. A UFRGS aponta que, assim como o repositório institucional, a ideia do repositório de dados será aceita pelas pró-reitorias e reitoria.

- e) desenvolvimento de políticas institucionais;

A comissão formada na UFSC está trabalhando a definição sobre as políticas institucionais que regem a gestão de dados de pesquisa na instituição. Na UFRGS, depois que o repositório de dados estiver delineado, a instituição passará a pensar nas políticas institucionais, com a participação de bibliotecários e pesquisadores de diversas áreas do conhecimento.

- f) cautela no desenvolvimento dos serviços;

As duas instituições demonstram cautela e cuidado no desenvolvimento de serviços. Além do levantamento das práticas e necessidades dos pesquisadores realizado pelas instituições, a UFSC está definindo as questões políticas. Começou com a questão da autoria dos dados, acesso aberto e todas as questões envolvidas com as políticas institucionais para a gestão de dados de pesquisa. A UFRGS já partiu para uma etapa mais prática, o desenvolvimento do repositório de dados. Contudo, demonstra cautela ao realizar estudo para identificar a ferramenta mais adequada e, assim, disponibilizar um protótipo do repositório. Acredita-se que as questões políticas entrarão em pauta.

g) colaboração e parcerias, internas e externas;

A UFSC formou uma comissão composta pela biblioteca, Pró-Reitoria de Pesquisa (PROPESQ), Pró-Reitoria de Pós-Graduação (PROPG), Superintendência de Governança Eletrônica e Tecnologia da Informação e Comunicação (SeTIC) e Secretaria de Inovação (SINOVA). Como atividades futuras, a instituição pensa em parcerias e colaboração com outras instituições externas como IBICT, RNP e Universidade Tecnológica de Munique. A UFRGS colabora, em parceria com a FURG, com um projeto da RNP e do IBICT. Além disso, a discussão e desenvolvimento das infraestruturas acontecem dentro do Centro de Processamento de Dados, que conta com bibliotecários, analistas e programadores.

h) definição da infraestrutura tecnológica;

A UFRGS realizou estudo para identificar e avaliar as principais ferramentas para o desenvolvimento de repositório de dados. A partir dessa avaliação criteriosa e desenvolvimento de um protótipo para testes será definida a infraestrutura que a universidade utilizará. A UFSC planeja estudar e definir a infraestrutura que a instituição utilizará, após a implementação das políticas.

i) criação de um projeto piloto;

Assim que a política institucional estiver definida, a UFSC planeja a criação de um projeto piloto com um ou dois grupos de pesquisa, antes de se começar a oferecer todo o conjunto de serviços e infraestrutura à instituição.

j) conscientização dos pesquisadores;

A UFSC realizou evento para a conscientização de sua comunidade cujo tema foi a gestão adequada dos dados de pesquisa. Diversas outras instituições interessadas na questão participaram. O evento também teve o caráter de formação de bibliotecários

e de pesquisadores. A UFRGS destaca a importância do convencimento do papel do pesquisador nas questões da gestão de dados. Uma ação de destaque realizada foi a entrevista com pesquisadores. Além de conhecer melhor suas práticas, promove a conscientização para a gestão de dados.

k) papel do bibliotecário;

Nas duas instituições, o bibliotecário é o principal articulador das ações voltadas para o desenvolvimento de serviços e infraestrutura para a gestão de dados de pesquisa. Na UFRGS, contudo, o bibliotecário, que incita as ações, não está diretamente ligado à biblioteca.

l) capacitação e formação das equipes;

Na UFSC, a ideia é que, após o estabelecimento das políticas e definição da infraestrutura tecnológica, seja realizada a capacitação da equipe para o suporte aos pesquisadores. A UFRGS não deixou explícita a ideia de capacitação, mas lembrou que eles tiveram dificuldade com a gestão de dados por ser este um assunto totalmente novo. Precisa, pois, compreender diversas questões, como formato de dados, e ferramentas. Sugere a importância da capacitação da equipe para essas novas questões.

m) definição das funções e responsabilidades;

A UFSC apontou a importância da definição dos fluxos, funções e responsabilidades dos diversos atores com a gestão de dados. Alguns serviços já são oferecidos por outras instâncias, como a Pró-reitoria de Pesquisa e a Secretaria de Inovação. A definição das funções e responsabilidades evitaria a duplicação de serviços. A biblioteca encaminharia as demandas para essas outras instâncias definidas.

n) adequação dos serviços e infraestruturas às demandas das áreas do conhecimento.

A UFRGS relata que se deve fornecer infraestruturas diferenciadas, dependendo da área do conhecimento. Algumas áreas podem não colocar os dados trabalhados nos repositórios, porque possuem repositórios específicos fora do país. Contudo, mesmo que não coloquem os dados trabalhados, essas áreas precisarão de um local para armazenamento dos dados intermediários ou dos dados brutos. Outras áreas precisarão de infraestrutura para depositar desde o estado bruto até o dado final.

A seção seguinte apresenta os desafios e dificuldades apontadas na literatura e nas duas entrevistas realizadas.

4.4. DESAFIOS E DIFICULDADES

A literatura aponta alguns desafios e dificuldades para o oferecimento dos SSGDPs pelas instituições. Algumas dessas dificuldades são apresentadas no Quadro 23.

QUADRO 23 – Dificuldades no oferecimento de SSGDPs

Dificuldades	AUTOR (ano)
Habilidades e conhecimentos limitados/ falta de especialização da equipe/qualificação e treinamento de pessoal	(GOLDMAN; KAFEL; MARTIN, 2015; FEARON <i>et al.</i> , 2013; CREAMER <i>et al.</i> , 2012; BROWN, 2010; LEWIS, 2010; SOEHNER; STEEVES; WARD, 2010; GOLD, 2007a)
Ausência de apoio institucional	(FEARON <i>et al.</i> , 2013; CREAMER <i>et al.</i> , 2012; RABOIN; REZNIK-ZELLEN; SALO, 2012; SOEHNER; STEEVES; WARD, 2010)
Ausência de financiamento e recursos – inclui a sustentabilidade dos serviços	(FEARON <i>et al.</i> , 2013; CREAMER <i>et al.</i> , 2012; BROWN, 2010; SOEHNER; STEEVES; WARD, 2010)
Ausência de conhecimento específico de domínio, compreensão das práticas de pesquisa, diversidade dos dados e cultura dos pesquisadores	(COX; VERBAAN; SEN, 2012; LEWIS, 2010; HENTY, 2008)
Ausência de uma política institucional	(GOLDMAN; KAFEL; MARTIN, 2015; FEARON <i>et al.</i> , 2013; SOEHNER; STEEVES; WARD, 2010)
Conscientização sobre a gestão de dados e envolvimento dos pesquisadores	(FEARON <i>et al.</i> , 2013; GOLDMAN; KAFEL; MARTIN, 2015; SOEHNER; STEEVES; WARD, 2010)
Falta de pessoal dedicado, falta de tempo da equipe, tamanho reduzido da equipe, problema com a captação de mais funcionários da biblioteca para as atividades de gestão de dados	(GOLDMAN; KAFEL; MARTIN, 2015; FEARON <i>et al.</i> , 2013; RABOIN; REZNIK-ZELLEN; SALO, 2012)
Infraestrutura tecnológica	(FEARON <i>et al.</i> , 2013; LEWIS, 2010; SOEHNER; STEEVES; WARD, 2010)
Pró-atividade junto aos pesquisadores	(LYON, 2012; COX; VERBAAN; SEN, 2012; HENTY, 2008)
Falta de <i>marketing</i> e conscientização sobre os serviços oferecidos pelas bibliotecas ou estratégias ineficazes de promoção dos serviços ou do papel da biblioteca	(GOLDMAN; KAFEL; MARTIN, 2015; FEARON <i>et al.</i> , 2013;)
Formação de parcerias e colaboração	(FEARON <i>et al.</i> , 2013; SOEHNER; STEEVES; WARD, 2010)
Ambiguidade dos requisitos das agências financiadoras	(FEARON <i>et al.</i> , 2013)
Ausência de experiência dos bibliotecários para trabalhar com pesquisa	(COX; VERBAAN; SEN, 2012)
Colaboração entre pesquisadores e departamentos	(GOLDMAN; KAFEL; MARTIN, 2015)
Compreensão correta da importância da gestão de dados	(BROWN, 2010)
Confusão em torno da definição de 'serviços de dados'	(GOLDMAN; KAFEL; MARTIN, 2015)
Continuidade do atendimento às necessidades dos pesquisadores na medida em que elas evoluem	(RABOIN; REZNIK-ZELLEN; SALO, 2012)
Definição do escopo dos serviços	(FEARON <i>et al.</i> , 2013)

Definição dos papéis da equipe	(FEARON <i>et al.</i> , 2013)
Educação e formação do corpo docente para a gestão de dados de pesquisa	(FEARON <i>et al.</i> , 2013)
Integração de serviços com estruturas de pessoal novas ou existentes	(RABOIN; REZNIK-ZELLEN; SALO, 2012)
Lutas territoriais entre os vários departamentos da instituição	(CREAMER <i>et al.</i> , 2012)
Mudanças na equipe e na administração	(GOLDMAN; KAFEL; MARTIN, 2015)
Não disposição dos pesquisadores para o compartilhamento de dados	(GOLDMAN; KAFEL; MARTIN, 2015)
Redução de outras prioridades da equipe em detrimento dos papéis relacionados à gestão de dados de pesquisa	(COX; VERBAAN; SEN, 2012)

Fonte: Elaborado pela autora.

Observamos que as dificuldades mais citadas na literatura dizem respeito à defasagem nas habilidades e conhecimento da equipe; falta de apoio institucional; falta de financiamento; desconhecimento do domínio (práticas, dados e cultura dos pesquisadores); ausência de política institucional; conscientização dos pesquisadores; tamanho, tempo e equipe dedicada às atividades de gestão de dados de pesquisa; infraestrutura tecnológica e; proatividade junto aos pesquisadores. Além dessas, o Quadro 23 aponta outras importantes dificuldades que não podem deixar de ser consideradas.

Essas dificuldades podem ser vistas como desafios a serem enfrentados para o oferecimento de um conjunto de SSGDPs, de modo que os serviços atendam às necessidades dos pesquisadores e que a biblioteca mantenha o seu protagonismo e relevância perante a comunidade.

Embora as duas instituições entrevistadas estejam em fase de desenvolvimento dos SSGDPs, depreendemos algumas dificuldades e desafios.

Na UFSC:

a) mudanças na equipe e na administração;

[...] a gente conseguiu apresentar a ideia. De início [...] a abertura foi boa. Só que, também, a gente teve muito percalço institucional [...] a UFSC passou por uns momentos bem complicados nesses últimos anos, troca de gestão [...] inesperadas [...] (P1, 2019).

[...] as coisas que aconteceram, acabaram trazendo transtorno bem grande institucionalmente, que refletiu em todos os órgãos, inclusive na biblioteca e nas iniciativas de todo mundo [...] (P1, 2019).

b) ausência de política institucional;

[...] a nossa instituição não tem políticas claras, nem para suas publicações, muito menos para dados [...] (P1, 2019).

Como é que a gente vai falar para o pesquisador. Olha, tu podes depositar os dados aqui no nosso repositório, mas ele vai questionar tudo isso! É aberto ou não é aberto? O que eu posso ter de privacidade, o que eu não posso? (P1, 2019).

Eles, [os pesquisadores] até nos questionários, não tinham certeza de quem eram os donos da informação (P1, 2019).

c) infraestrutura tecnológica;

[...] a infraestrutura, qual repositório a gente vai usar? Vai ser o que a gente já tem? A gente usa o Dspace. Ou vamos ter um separado e outro *software*, inclusive, para dados? (P1, 2019).

d) formação de parcerias e colaboração;

Vamos fazer parceria com esta instituição da Alemanha, por exemplo? (P1, 2019).

e) habilidades e conhecimentos limitados/ falta de especialização da equipe/qualificação e treinamento de pessoal;

[...] acho que é muito importante a capacitação das equipes, porque, quem que vai dar esse suporte, né? (P1, 2019).

f) definição dos papéis da equipe.

O que vai ser função da biblioteca? O que vai ser função de outros órgãos? O que, às vezes, já existe? Por exemplo, a pró-reitoria de pesquisa já tem alguns serviços específicos. Então, tem coisas que a gente deve, pelo menos, saber direcionar (P1, 2019).

Na UFRGS:

a) ausência de política institucional;

Assim que a gente tiver delineado o repositório, feito testes, teremos que pensar na política (P2, 2019).

b) conscientização sobre a gestão de dados e envolvimento dos pesquisadores;

[...] nós vamos ter que chamar bibliotecários, pesquisadores de diversas áreas e começar a fazer o mesmo processo que fizemos com o repositório institucional. Envolver as pessoas para que essas pessoas repliquem isso e o repositório funcione (P2, 2019).

[...] e acho que esse vai ser o maior problema com os pesquisadores, deles organizarem já, desde o plano de gestão de dados já organizarem a pesquisa, até o dado chegar no repositório (P2, 2019).

c) habilidades e conhecimentos limitados/ falta de especialização da equipe/qualificação e treinamento de pessoal;

A dificuldade é a gente conhecer esse mundo totalmente diferente (P2, 2019).

[...] o que são formatos? O que são formatos de dados? A gente entende de formato Word, PDF e tal, mas tem milhares de formatos. Tem milhares de ferramentas que [podem] ser incorporadas aos dados para que [...] possam ser acessíveis (P2, 2019).

[...] os bibliotecários também, nas setoriais, vão ser treinados para que possam dar orientações, tanto no plano de gestão de dados, como em que repositório, como depositar os dados, como levantar os dados (P2, 2019).

- d) ausência de conhecimento específico de domínio, compreensão das práticas de pesquisa, diversidade dos dados e cultura dos pesquisadores.

A gente entrevistou um professor da biologia molecular. Ele deu uma aula de repositórios de dados científicos, porque essa é uma área do conhecimento que já está acostumada com isso. Tem áreas que vão precisar de infraestrutura intermediária e não o dado final [...] Em outras áreas eles não têm nada. Vão precisar desde [colocar no repositório] o dado bruto até o dado final, o que vai para o artigo [...] (P2, 2019).

Algumas vezes, as dificuldades listadas para as duas instituições entrevistadas foram colocadas em forma de dúvidas que precisavam ser resolvidas para o andamento do oferecimento dos serviços. Assim, mesmo que a fala dos entrevistados não apontem diretamente as dificuldades, elas foram percebidas por meio das dúvidas externadas ou pelos questionamentos de como iriam agir diante de determinadas situações.

4.5. AÇÕES PARA VIABILIZAÇÃO E MANUTENÇÃO

Para o desenvolvimento, funcionamento e manutenção de SSGDPs nas instituições, são necessárias algumas ações. Essas ações foram chamadas neste trabalho de ações para viabilização de serviços e ações para manutenção de serviços.

As ações para viabilização são intervenções que devem ser realizadas antes de os serviços entrarem em funcionamento. Muitas vezes, é a partir delas que ocorre o desenvolvimento dos serviços ou infraestruturas para o suporte à gestão de dados de pesquisa gerados na instituição.

Outras ações devem ser constantes durante todo o tempo, visando garantir que os serviços estejam em pleno funcionamento ao longo do tempo. Essas são as ações para manutenção de serviços.

O Quadro 24 apresenta as ações para viabilização dos serviços que ocorrem para que possam ser construídos, ou entrem em funcionamento, e as ações para manutenção que devem ocorrer durante todo o tempo, incluindo a fase de elaboração/desenvolvimento dos serviços e a fase pós-implementação de serviços.

QUADRO 24 - Ações para viabilização e manutenção de SSGDPs nas instituições

Ações		Definição	Viabil	Manut
1	Criação de grupos de trabalho, comissão ou força-tarefa	Criação de grupos de trabalho, comissão ou força-tarefa para discutir e operacionalizar a gestão e curadoria de dados na instituição. Inclui a discussão sobre a gestão e curadoria de dados, de forma geral e sobre os serviços que podem ser oferecidos, atendendo às necessidades dos pesquisadores.	X	
2	Previsão de custos	Previsão detalhada dos custos para gestão e curadoria dos dados produzidos na instituição. Essa previsão deve abordar os custos para manutenção de dados seguros e acessíveis ao longo do tempo, bem como infraestrutura tecnológica, trabalho da equipe, contratação de pessoal e capacitação da equipe de trabalho.	X	
3	Aprovação e apoio gerencial	Aprovação da alta cúpula da instituição. É necessário convencer a gestão da instituição de que os dados de pesquisa precisam ser gerenciados adequadamente. A adesão da gestão irá determinar a taxa de sucesso na implementação dos serviços.	X	X
4	Desenvolvimento e implementação de políticas e diretrizes	As definições políticas perpassam questões como o uso de padrões de dados e metadados estabelecidos, acesso e compartilhamento dos dados, retenção e preservação de dados, incluindo a possibilidade de depósito em uma infraestrutura local ou em infraestruturas externas. Os critérios para seleção dos dados que serão mantidos são baseados na capacidade de gestão e no valor do conjunto de dados - origem observacional, computacional ou experimental - o que influencia nas decisões de curadoria desses dados), exigência de plano de gestão de dados, privacidade, tratamento dos dados sensíveis, propriedade intelectual, propriedade/autoria dos dados, adesão aos princípios de gestão de dados fortemente estabelecidos, o escopo dos dados cobertos pela política, papéis e responsabilidades na gestão de dados em todas as suas fases, execução e monitoramento da política. As políticas e diretrizes podem englobar, dentre outras coisas: a obrigatoriedade da gestão adequada e compartilhamento de dados para a obtenção de financiamento para a pesquisa na instituição, se for o caso; e a determinação de cotas de armazenamento nos repositórios.	X	X
5	Planejamento estratégico	Elaborar planejamento estratégico para o desenvolvimento de serviços na instituição. Um bom plano estratégico representa apenas o início do trabalho real de desenvolvimento de serviços. Define medidas de resultados e garante que todos os principais participantes estejam se movendo na mesma direção.	X	X
6	Definição do escopo dos serviços	Definição do escopo dos serviços, seus objetivos, público, limite de operações <i>etc.</i> , tendo em vista a missão e as prioridades institucionais.	X	
7	Mapeamento das atividades e fluxos de trabalho	Mapeamento das atividades e fluxos de trabalho com base no ciclo de vida dos dados, desde o momento do planejamento, até a disponibilização para reuso. Os serviços serão desenvolvidos a partir desse mapeamento.	X	

8	Identificação de colaboradores e possíveis patrocinadores e estabelecimento parcerias	Ações referentes à identificação de colaboradores e estabelecimento parcerias para oferecimento de serviços de gestão e curadoria de dados.	X	X
9	Identificação das percepções, práticas e necessidades dos pesquisadores	Identificação das percepções, práticas e necessidades dos pesquisadores em relação à gestão e curadoria de dados. Os serviços serão desenvolvidos a partir desse levantamento.	X	
10	Identificação de recursos e serviços disponíveis na instituição	Identificação dos recursos e serviços relacionados à gestão e curadoria de dados existentes na instituição. Inclui a identificação de lacunas nos serviços disponíveis.	X	
11	Identificação das habilidades e conhecimentos necessários aos profissionais	Identificação das habilidades e conhecimentos necessários aos profissionais para atender às necessidades de gestão e curadoria de dados dos pesquisadores.	X	X
12	Definição e atribuição das funções e responsabilidades	Definição e criação das funções e responsabilidades no suporte à gestão de dados que serão realizadas pelas bibliotecas e pelos demais parceiros envolvidos. As funções devem ser atribuídas às equipes ou pessoas. Esse processo envolve a redefinição das funções já realizadas e reestruturação das equipes.	X	
13	Capacitação/ desenvolvimento da equipe	Deve-se promover a capacitação contínua da equipe para assumir novas funções inerentes à gestão de dados. São elas: formato de dados de pesquisa, <i>softwares</i> , padrões de metadados, indexação de dados. É necessária a capacitação para lidar com consultas relacionadas à gestão de dados durante todo o ciclo de vida.	X	X
14	Adoção de boas práticas ou diretrizes de outras instituições	Adoção das melhores práticas estabelecidas em outras instituições que já possuem a <i>expertise</i> na gestão e curadoria de dados.	X	X
15	Requisitos éticos em pesquisa	Atuar junto ao comitê de ética em pesquisa ou órgão semelhante para documentar e explicar as etapas, atividades e requisitos de gestão de dados necessários para coletar e trabalhar com dados de forma ética.	X	X
16	Definição da infraestrutura tecnológica para armazenamento de dados	Identificação das possíveis opções para armazenamento de dados. As opções podem ser repositórios de dados locais ou repositórios externos, multidisciplinares ou específicos por disciplina. É importante que a opção escolhida siga os princípios FAIR (<i>Findable, Accessible, Interoperable, Reusable</i>). Algumas opções são: Dataverse, DSpace, CKAN, Figshare, Dryad, Zenodo, DataHub, DANS.	X	

17	Seleção de padrão de dados e metadados	É necessário usar formatos de dados padronizados, permitindo que eles sejam conversíveis, lidos em qualquer lugar e expansíveis. Além disso, é essencial a definição de qual padrão de metadados utilizar, tendo em vista a completude dos metadados e o custo x benefício. Os padrões de metadados devem definir os elementos e as estruturas usadas para descrever os conjuntos de dados (coleta - quem, onde, como e quais tipos; as condições de acesso e os termos de uso do conjunto de dados, dentre outras coisas), fornecendo informações suficientes para entender e usá-los, essenciais para sua descoberta e uso efetivo. O ideal é que a seleção dos padrões seja realizada juntamente com os pesquisadores, pois eles conhecem melhor os seus dados. Podem-se criar perfis de aplicação, não sendo necessário usar todos os elementos do padrão escolhido.	X	
18	Desenvolvimento de serviços, ferramentas, infraestruturas	Desenvolvimento de ferramentas (ex.: criação do PGD <i>online</i> , formulário para depósito de dados), repositórios de dados, serviços, plataformas.	X	X
19	Desenvolvimento de padrões	Criação de padrões relacionados à publicação, descrição, citação, descoberta e reutilização de dados.	X	X
20	Elaboração de documentos de orientação	Guias, manuais, <i>templates</i> envolvendo todos os aspectos da gestão e curadoria de dados.	X	X
21	Criação de um projeto piloto	Criação de um projeto piloto que abranja apenas algumas áreas ou projetos de pesquisa antes da implementação dos serviços em toda a instituição. O projeto piloto é um importante laboratório, onde os serviços e infraestruturas podem ser testadas e melhoradas.	X	
22	Desenvolvimento de <i>website</i>	Desenvolvimento de uma página na <i>web</i> para informar e conectar pesquisadores às informações básicas sobre todos os aspectos da gestão de dados, além de conscientizar sobre a gestão de dados e seus requisitos. O <i>website</i> deve reunir, organizar e apresentar os principais recursos ou serviços disponíveis para os pesquisadores.	X	X
23	Divulgação, comunicação e promoção dos serviços	Persuasão e envolvimento dos pesquisadores e toda a comunidade universitária sobre a importância e os benefícios da gestão de dados de pesquisa. Realização de eventos (<i>workshops</i> , palestras, apresentações em reuniões com o corpo docente, pesquisadores), boletins informativos, <i>marketing</i> dos serviços <i>etc.</i>	X	X
24	Avaliação dos serviços	Avaliação contínua dos serviços de forma a detectar e corrigir as falhas, identificar pontos fracos e manter os serviços relevantes para os usuários e instituição. A avaliação objetiva dos serviços pode incluir uma variedade de métricas como a quantidade de depósitos de dados, número de <i>workshops</i> ou sessões de treinamentos ministrados, colaborações externas, uso do <i>website etc.</i>		X

Fonte: Elaborado pela autora.

As ações apresentadas no Quadro 24 não são obrigatórias, mas sugestões para o desenvolvimento de serviços relevantes, bem como sua manutenção e sustentabilidade ao

longo do tempo, de forma a atender as necessidades dos pesquisadores em relação à gestão dos dados gerados em suas pesquisas.

Essas ações podem servir de base para o planejamento estratégico da instituição ou da biblioteca, auxiliando gestores a: analisar o cenário da instituição; refletir e determinar os caminhos que a instituição deve seguir e os serviços que ela pode oferecer; traçar estratégias e metas; planejar ações, recursos e decisões; e operacionalizar a criação e/ou inclusão de serviços de forma a alcançar seus objetivos e metas da instituição.

O capítulo seguinte apresenta as considerações finais deste trabalho, incluindo as contribuições da pesquisa, limitações e estudos futuros.

5. CONSIDERAÇÕES FINAIS

Atualmente, os dados de pesquisa ganharam destaque no cenário mundial, principalmente a partir do momento em que as pesquisas científicas intensificaram-se tanto no que se refere à geração como no uso de dados, fenômeno conhecido como *eScience*.

Esse fenômeno é observado com mais intensidade nas áreas da chamada *big science*, que é marcada por grandes projetos científicos espalhados pelo mundo. Esses projetos usam e produzem enormes quantidades de dados, que são altamente estruturados e possuem práticas já consagradas de compartilhamento e reuso.

Os pequenos conjuntos de dados, produzidos principalmente no âmbito das universidades, na chamada cauda longa da ciência, ainda constituem gargalos em termos de gestão, preservação, compartilhamento e reuso. As pesquisas na cauda longa da ciência são conduzidas em menor escala, por indivíduos ou pequenos grupos. Os dados gerados por essas pesquisas são altamente heterogêneos, geralmente insubstituíveis, caros de reproduzir e raramente reutilizados, ficando isolados em escritórios e laboratórios dos pesquisadores.

A gestão adequada dos dados de pesquisa, de modo que eles possam ser preservados e reutilizados, esbarra em questões como falta de recursos, infraestrutura, conhecimento, padronização, políticas, dentre outras questões.

As bibliotecas apresentam enorme potencial para encarar esse desafio nas instituições. Elas já estão inseridas no ciclo de comunicação, principalmente na organização e fornecimento de informações e no oferecimento de serviços de formação e orientação aos seus usuários.

Quanto à atuação bibliotecário nesse contexto, parece que ele deve ser uma espécie de 'super bibliotecário'. Novas questões lhe são propostas. Surgem novos papéis para atuação. O fornecimento de serviços relacionados à gestão de dados de pesquisa seria, contudo, uma extensão do papel tradicional das bibliotecas e bibliotecários, que possuem longa experiência na gestão e acesso à informação. A gestão de dados de pesquisa demanda mudança ou ampliação do foco dessas instituições, transpondo a gestão de coleções para uma participação mais ativa no processo de pesquisa.

Os serviços que as bibliotecas podem oferecer no âmbito da gestão de dados de pesquisa são chamados nesta tese de Serviços de Suporte à Gestão de Dados de Pesquisa (SSGDPs). Tenopir, Birch e Allard (2012) denominam esses serviços como aqueles que as bibliotecas oferecem à gestão de dados de pesquisa, abordando todo o seu ciclo de vida, desde o momento do planejamento da pesquisa até sua reutilização.

Dito isso, embora os caminhos para o desenvolvimento de SSGDPs nas instituições sejam únicos, eles envolvem questões que perpassam a qualificação da equipe, desenvolvimento de políticas, infraestruturas e serviços, orientados pelas necessidades dos usuários e pelo ciclo de vida dos dados, além da compreensão do contexto da própria instituição.

Este trabalho teve como objetivo deste estudo é propor uma taxonomia de serviços voltados à gestão de dados de pesquisa gerados nas universidades, bem como sintetizar dificuldades, desafios e ações necessárias para a viabilização e manutenção desses serviços nas instituições.

Esta pesquisa apresenta cinco produtos: 1) taxonomia de serviços; 2) compilação dos SSGDPs; 3) contextualização dos esforços para o desenvolvimento de SSGDPs nas instituições entrevistadas; 4) síntese dos desafios e dificuldades; 5) síntese das ações para viabilização e manutenção dos serviços.

Quanto à taxonomia de serviços, embora existam outras classificações de serviços como apresentados pelos trabalhos de Tenopir *et al.* (2014, 2015, 2017) que dividem os serviços em consultivos/informacionais e serviços técnicos e Reznik-Zellen; Adamick e McGinty (2012) que dividiram os serviços em Educacionais, Consulta ou Infraestrutura, optamos por criar uma nova classificação mais adequada ao que foi encontrado na literatura, tomando como base os trabalhos desses autores.

Assim, a taxonomia proposta foi estruturada em três categorias principais, Infraestrutura, Suporte e Educacional.

A categoria Infraestrutura, criada com base na subdivisão proposta por Reznik-Zellen; Adamick e McGinty (2012), inclui disponibilização de ferramentas, métodos, *softwares* e plataformas para a gestão e curadoria de dados de pesquisa, incluindo os repositórios de dados, além de outras infraestruturas como políticas institucionais – indicadas por diversos autores como importante infraestrutura para a gestão e curadoria de dados -, ferramentas para análise de redes sociais, assistência à produtividade e trabalho colaborativo.

A categoria Suporte representa a união das categorias serviços consultivos/informacionais e serviços técnicos proposta por Tenopir *et al.* (2014, 2015, 2017), uma vez que o *corpus* analisado não permitiu identificar o que era uma atividade realizada pela biblioteca e o que era uma atividade de suporte. Dessa forma, a categoria é formada pelos serviços técnicos (1) e pelas atividades de suporte (2).

Os serviços técnicos (1), tipo de assistência técnica, envolve os serviços realizados pelas bibliotecas como: documentação de dados e metadados; depósito de dados; preservação; garantia de acesso; publicação, dentre outros.

Essa categoria engloba também atividade de suporte (2) como informação, orientação, aconselhamento e consultoria, de modo presencial ou *online*, sobre os mais variados assuntos referentes à gestão e curadoria de dados, como a elaboração de planos de gestão de dados, padrões de metadados, acesso aberto aos dados, melhores práticas de pesquisa, políticas de dados, serviços oferecidos e outros.

A categoria Educacional, criada com base na subdivisão proposta por Reznik-Zellen; Adamick e McGinty (2012), abrange os serviços educacionais voltados para a formação em gestão e curadoria de dados realizados pelas bibliotecas como *advocacy*, conscientização, incentivo em questões de gestão e curadoria de dados. Também envolve a conscientização sobre os serviços oferecidos pelas bibliotecas; eventos como conferências, seminários, simpósios, palestras e painéis relacionados à gestão e curadoria de dados e outros eventos relacionados à identificação das necessidades dos pesquisadores, políticas, reprodutibilidade da pesquisa e serviços oferecidos; elaboração e disponibilização de guias, *templates*, tutoriais ou *help* relacionados à gestão e curadoria de dados.

A categoria Educacional considera, ainda, questões relacionadas às políticas e tecnologias, infraestruturas e ferramentas; treinamentos, oficinas, cursos, aulas relacionadas à gestão e curadoria de dados, bem como outros tipos de treinamentos como àqueles relacionados às questões de políticas, serviços, tecnologias, infraestruturas e ferramentas oferecidas.

As três categorias principais (Infraestrutura, Suporte e Educacional) são subdivididas nas subcategorias 'Gestão e curadoria de dados' e 'Outros'. Essas duas subcategorias têm subdivisões adicionais que são replicadas para cada uma das categorias principais.

Na subcategoria 'Gestão e curadoria de dados' temos as seguintes subdivisões: 000 Gestão e curadoria geral; 001 Planejamento; 002 Obtenção; 003 Organização e formato; 004 Documentação de dados e metadados; 005 Avaliação e seleção; 006 Armazenamento; 007 *Copyright*, propriedade intelectual e licenciamento; 008 Questões éticas, legais e de consentimento; 009 Publicação e citação; 010 Uso e reuso; e 011 Mineração, análise e visualização.

A subcategoria 'Outros' é dividida em: 012 Análise de redes sociais; 013 Assistências à produtividade e trabalho colaborativo; 014 Concepção e gestão de bases de dados; 015 Desenvolvimento de diretrizes, padrões, serviços e infraestruturas; 016 Desenvolvimento de disciplina ou *curriculum*; 017 Gestão de pesquisa e comunicação científica; 018 Identificação

das percepções, práticas e necessidades; 019 Participação em projeto ou processo de pesquisa; 020 Políticas; 021 Rastreamento de impacto; 022 Reprodutibilidade da pesquisa; 023 Serviços; e 024 Tecnologias, infraestruturas e ferramentas.

Com base nas menções aos serviços identificadas nos artigos 165 artigos e em duas entrevistas, foi possível identificar a categoria Suporte como a mais frequentemente mencionada. Os serviços categorizados como Suporte foram os assíduos, com 1612 citações, o que corresponde 59,03% das referências aos serviços encontradas nos 167 documentos (165 artigos + duas entrevistas) utilizados para a construção da taxonomia.

Em seguida despontaram os serviços da categoria Educacional, com 826 referências, representando 30,24% das citações. Por último, os serviços da categoria Infraestrutura apareceram em 293 referências, representando 10,73% do total das citações aos serviços encontradas.

As duas categorias mais presentes – Suporte e Educacional – compreendem os serviços de referência, instrução e orientação, representando uma extensão dos serviços tradicionais já realizados pelas bibliotecas, com o foco, agora, nos dados de pesquisa.

Na categoria Infraestrutura, subcategoria ‘Gestão e curadoria de dados’ sobressaíram as infraestruturas para armazenamento de dados ‘INF1-006 Armazenamento’. Na subcategoria ‘Outros’ sobressaíram as menções às políticas institucionais.

Na categoria Suporte, subcategoria ‘Gestão e curadoria de dados’ houve destaque para o suporte ao armazenamento de dados ‘SUP1-006 Armazenamento’, seguido pela subdivisão ‘SUP1-001 Planejamento’. Nesta análise, foi desconsiderada a subdivisão ‘SUP1-000 Gestão e curadoria – geral’ devido à sua generalidade. Por outro lado, a subdivisão menos presente foi ‘SUP1-008 Questões éticas, legais e de consentimento’. Na subcategoria ‘Outros’ destacaram-se o suporte com tecnologias, infraestruturas e ferramentas, ‘SUP2-024 Tecnologias, infraestruturas e ferramentas’.

Na categoria Educacional, subcategoria ‘Gestão e curadoria de dados’, excetuando a subdivisão ‘EDU1-000 Gestão e curadoria – geral’, por ser apresentada de forma ampla, generalista sem especificar exatamente os serviços, os serviços mais mencionados foram aqueles voltados para o planejamento, ‘EDU1-001 Planejamento’, que inclui, principalmente, o Plano de Gestão de Dados (PGD). Na subcategoria ‘Outros’, sobressaíram as menções aos serviços categorizados em ‘EDU2-024 Tecnologias, infraestruturas e ferramentas’.

Somando-se todas as citações aos serviços, independentemente da categoria principal, observamos que, na subcategoria ‘Gestão e curadoria de dados’, sobressaíram os serviços

de armazenamento de dados '006 Armazenamento', seguido por '001 Planejamento' e '004 Documentação de dados e metadados'. Na subcategoria 'Outros' destacaram-se as subdivisões '024 Tecnologias, infraestruturas e ferramentas' e '020 Políticas'.

A estruturação da taxonomia pode ser conferida no Apêndice H, enquanto a quantificação das citações aos serviços em cada uma das categorias e subcategorias pode ser consultada no Apêndice I. A lista completa de serviços em cada uma das categorias, outro produto desta tese, pode ser consultada nos apêndices J, K e L.

Como apresentado anteriormente, a taxonomia proposta e a compilação de serviços foram realizadas a partir da literatura e das entrevistas.

As entrevistas, tomadas individualmente, permitiram contextualizar os esforços relacionados à gestão de dados de pesquisa nas duas instituições entrevistadas, Universidade Federal de Santa Catarina (UFSC) e Universidade Federal do Rio Grande do Sul (UFRGS). A contextualização dos serviços nessas instituições é produto deste trabalho. Essas instituições encontram-se em fase de desenvolvimento de serviços ou infraestruturas.

Embora as duas instituições estejam em fase de desenvolvimento de serviços e infraestruturas, os passos tomados por cada uma delas diferenciaram-se. A UFSC iniciou os trabalhos liderada pela biblioteca, formando uma comissão com a participação de diversas instâncias como a Pró-reitoria de Pesquisa, a Secretaria de Inovação e a Superintendência de Governança Eletrônica e Tecnologia da Informação e Comunicação. Uma das primeiras ações da comissão foi a realização de um levantamento das práticas e necessidades dos pesquisadores da instituição. Em seguida, foi realizado um evento para conscientização de sua comunidade - I Seminário de Suporte à Pesquisa e Gestão de Dados Científicos. No momento da entrevista, realizada em abril de 2019, a comissão estava se dedicando à elaboração da política institucional para a gestão de dados de pesquisa. Essa questão foi considerada pela entrevistada como uma das principais, que deveria ser realizada antes mesmo de oferecer algum tipo de serviço ou infraestrutura para a gestão de dados de pesquisa.

A UFRGS possui duas frentes de trabalho em relação à gestão de dados de pesquisa. Na primeira frente, a Universidade colabora em um projeto mantido pela Rede Nacional de Ensino e Pesquisa (RNP) e pelo Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT), com a participação ainda da Universidade Federal do Rio Grande (FURG). O segundo projeto da UFRGS é a criação de um repositório de dados. Os dois projetos são liderados na instituição por uma bibliotecária que atualmente atua como docente. Apesar da participação bibliotecária, os projetos não são mantidos pela biblioteca, mas pelo Centro de

Processamento de Dados da instituição. Na UFRGS, após o levantamento das práticas e necessidades dos pesquisadores, em nível nacional, no projeto da RNP e IBICT, foi realizada a identificação e análise das ferramentas para repositório de dados. Essas informações são compartilhadas nos dois projetos. Assim, atualmente, estão trabalhando com um protótipo de repositório utilizando o Dataverse. Após a definição da infraestrutura para armazenamento dos dados, irão se dedicar a outras questões como as políticas.

A literatura e as entrevistas permitiram a identificação dos desafios e dificuldades no desenvolvimento de SSGDPs, o quarto produto desta tese; e elencar uma série de ações para viabilização e manutenção dos serviços na instituição, considerada como o quinto produto da tese.

As dificuldades tornam-se desafios para as bibliotecas e instituições que desejam se arriscar na área da gestão de dados de pesquisa. As principais dificuldades dizem respeito à defasagem nas habilidades e conhecimento da equipe; falta de apoio institucional; falta de financiamento; desconhecimento do domínio (práticas, dados e cultura dos pesquisadores); ausência de política institucional; conscientização dos pesquisadores; tamanho, tempo e equipe dedicada às atividades de gestão de dados de pesquisa; infraestrutura tecnológica; e proatividade junto aos pesquisadores.

As ações para viabilização ocorrem para que os serviços possam ser construídos ou para que entrem em funcionamento. As ações para manutenção ocorrem durante todo o tempo, incluindo a fase de elaboração/desenvolvimento dos serviços e a fase pós-implementação de serviços.

Podemos citar algumas ações importantes como a criação de políticas para a gestão de dados; o apoio institucional; a formação de parcerias; a identificação das práticas e necessidades dos pesquisadores; a capacitação da equipe; a definição da infraestrutura tecnológica e dos padrões utilizados; a criação de documentação para orientação da comunidade; a conscientização dos pesquisadores; e a divulgação dos serviços e avaliação constante.

Consideramos que a categorização é trabalho altamente subjetivo, ainda que amparado na literatura. A taxonomia de serviços apresentada na tese passou por diversas reanálises e reagrupamentos a partir de um melhor entendimento da temática de pesquisa. Destacamos que terceiros podem ter opiniões divergentes em torno da categorização proposta. Dessa forma, a subjetividade e divergência na categorização pode ser considerada limitação deste estudo. Outra limitação consiste no fato de a análise estatística descritiva básica ter sido construída em cima das categorias e subcategorias, que, por sua vez, são formadas por

diversos serviços, não sendo possível, por exemplo, estimar numericamente cada um desses serviços.

Consideramos que a pesquisa cumpriu os objetivos propostos de identificar serviços, ações, dificuldades e desafios; contextualizar o oferecimento de serviços e infraestruturas nas universidades entrevistadas; sistematizar e propor uma taxonomia de serviços; sintetizar as dificuldades e desafios; e sintetizar as ações para viabilização e manutenção dos serviços nas instituições.

5.1. CONTRIBUIÇÕES DA PESQUISA

A importância da pesquisa apresenta-se por diversas razões:

- a) pode auxiliar os diversos atores envolvidos (instituição, bibliotecas, pesquisadores) no entendimento dos seus papéis na gestão dos dados de pesquisa;
- b) pode contribuir para a definição de uma política de dados para as universidades;
- c) pode servir como material de referência nos assuntos abordados no referencial teórico;
- d) pode servir como material de referência para o desenvolvimento de serviços, uma vez que o trabalho permitiu a identificação e sistematização dos principais serviços desenvolvidos pelas bibliotecas;
- e) pode servir como diretriz para instituições ou profissionais que desejam desenvolver SSGDPs ou adicionar serviços àqueles já desenvolvidos em suas instituições;
- f) pode encorajar as bibliotecas a ampliarem seus serviços oferecidos;
- g) pode encorajar os bibliotecários alargarem seu campo de atuação;
- h) pode promover debates a respeito da formação do profissional bibliotecário frente às necessidades surgidas a partir de um novo cenário para a comunicação científica;
- i) pode servir como base para o planejamento estratégico da instituição ou biblioteca que deseja implantar ou ampliar serviços.

5.2. ESTUDOS FUTUROS

Com estudos futuros, podem-se identificar os serviços disponibilizados pelas bibliotecas universitárias brasileiras bem como os desafios para sua implementação. Os serviços

elencados nesta tese foram majoritariamente obtidos pela revisão de literatura em contexto mundial, que muitas vezes não condiz com a realidade brasileira.

Apesar de o trabalho apresentar uma série de serviços que podem ser desenvolvidos nas instituições, cada uma deve, ao iniciar o desenvolvimento de SSGDPs, tomar como base as necessidades dos seus pesquisadores. Assim, sugere-se para estudos futuro a identificação dos fluxos de pesquisa nos laboratórios e grupos de pesquisa e o levantamento das necessidades de gestão e curadoria de dados apresentadas por esses pequenos laboratórios e grupos.

De forma geral, estudos sobre ferramentas para repositórios de dados poderiam ser realizados, assim como: políticas institucionais que abordam a gestão de dados; questões legais e éticas em relação aos dados de pesquisa; compartilhamento e reuso de dados; perfis e papéis dos profissionais envolvidos na gestão de dados (bibliotecários, arquivistas, analistas de sistemas e outros); interdisciplinaridade na gestão de dados; novas formatos ou tendências da comunicação científica incluindo os *data journals* e *data papers*. Além disso, todos os assuntos abordados no referencial teórico desta tese têm potencial para estudos mais aprofundados.

REFERÊNCIAS

- ABBOTT, D. *What is digital curation?* Edinburgh: DCC, 2010. Disponível em: <https://web.archive.org/web/20040607065214/http://www.dcc.ac.uk/what.html>. Acesso em: 2 out. 2017.
- AKERS, K. G. *et al.* Building support for research data management: biographies of eight research universities. *International Journal of Digital Curation*, v. 9, n. 2, p. 171–191, 2014. Disponível em: <http://www.ijdc.net/index.php/ijdc/article/view/327>. Acesso em: 27 ago. 2018.
- ALMEIDA, F.G.; CENDÓN, B.V. Serviços de descoberta: panorama nas bibliotecas das instituições públicas de ensino superior brasileiras. *Em Questão*, v. 25, n.2, p. 300-325, 2019. Disponível em: <https://seer.ufrgs.br/EmQuestao/article/view/81969/51976>. Acesso em: 01 ago. 2019.
- ASHLEY, K. Research data and libraries: who does what. *Insights: the UKSG journal*, v. 25, n. 2, p. 155–157, 2012. Disponível em: <http://insights.uksg.org/articles/10.1629/2048-7754.25.2.155>. Acesso em: 21 set. 2018.
- ASSOCIATION OF RESEARCH LIBRARIES (ARL). *Agenda for developing e-Science in research libraries*: final report and recommendations to the Scholarly Communication Steering Committee, the Public Policies Affecting Research Libraries Steering Committee, and the Research, Teaching, and Learning Steeri. Washington, D.C.: ARL, 2007. Disponível em: http://www.arl.org/bm~doc/ARL_EScience_final.pdf. Acesso em: 2 jul. 2017.
- AUCKLAND, M. *Re-skilling for research*: an investigation into the role and skills of subject and liaison librarians required to effectively support the evolving information needs of researchers. London: Research Libraries UK, 2012. Disponível em: <http://www.rluk.ac.uk/content/re-skilling-research>. Acesso em: 25 ago. 2017.
- BALL, A. *Review of data management lifecycle models*. Bath: University of Bath, 2012. Disponível em: <https://purehost.bath.ac.uk/ws/portalfiles/portal/206543/redm1rep120110ab10.pdf>. Acesso em: 19 jul. 2018.
- BARBROW, S.; BRUSH, D.; GOLDMAN, J. Research data management and services: resources for novice data librarians. *College & Research Libraries News*, v. 78, n. 5, p. 274, 2017. Disponível em: <http://crln.acrl.org/index.php/crlnews/article/view/16660>. Acesso em: 14 fev. 2019.
- BARDIN, L. *Análise de conteúdo*. Lisboa: Edições 70, 1977.
- BARDYN, T. *et al.* Health sciences libraries advancing collaborative clinical research data management in universities. *Journal of eScience Librarianship*, v. 7, n. 2, p. e1130 1-14, 2018. Disponível em: <https://escholarship.umassmed.edu/jeslib/vol7/iss2/4/>. Acesso em: 30 jul. 2018.
- BEAGRIE, N. Digital curation for science, digital libraries, and individuals. *The International Journal of Digital Curation*, v. 1, n. 1, p. 3–16, 2006. Disponível em: <http://www.ijdc.net/index.php/ijdc/article/view/6/2>. Acesso em: 14 set. 2017.
- BERMAN, E. Transforming information literacy in the Sciences through the lens of e-Science. *Communications in Information Literacy*, v. 7, n. 2, p. 161–170, 2013.
- BERMAN, E. An exploratory sequential mixed methods approach to understanding researchers' data management practices at UVM: findings from the qualitative phase. *Journal of eScience Librarianship*, v. 6, n. 1, p. e1097 1-22, 2017a. Disponível em: <http://escholarship.umassmed.edu/jeslib/vol6/iss1/7/>. Acesso em: 8 out. 2018.

BERMAN, E. An Exploratory Sequential Mixed Methods Approach to Understanding Researchers' Data Management Practices at UVM: Findings from the Quantitative Phase. *Journal of eScience Librarianship*, v. 6, n. 1, p. e1098 1-24, 2017b. Disponível em: < <http://escholarship.umassmed.edu/jeslib/vol6/iss1/6/>. Acesso em: 08 out. 2018.

BERMAN, E. An exploratory sequential mixed methods approach to understanding researchers' data management practices at UVM: integrated findings to develop research data services. *Journal of eScience Librarianship*, v. 6, n. 1, p. e1104 1-24, 2017c. Disponível em: <http://escholarship.umassmed.edu/jeslib/vol6/iss1/7/>. Acesso em: 22 ago. 2018.

BICARREGUI, J. *et al.* Data management and preservation planning for big science. *International Journal of Digital Curation*, v. 8, n. 1, p. 29–41, 14 jun. 2013. Disponível em: <http://www.ijdc.net/index.php/ijdc/article/view/247>. Acesso em: 22 fev. 2018.

BORGMAN, C. L. Research Data: Who Will Share What, with Whom, When, and Why? 2010, Beijing: [s.n.], 2010. p. 1–21. Disponível em: <http://works.bepress.com/borgman/238/>. Acesso em: 22 set. 2017.

BORGMAN, C. L. The conundrum of sharing research data. *Journal of the American Society for Information Science and Technology*, v. 63, n. 6, p. 1059–1078, 2012. Disponível em: <http://onlinelibrary.wiley.com/doi/10.1002/asi.22634/full>. Acesso em: 31 ago. 2017.

BORGMAN, C. L. *Big data, little data, no data: scholarship in the networked world*. Cambridge, USA; London: MIT Press, 2015.

BORGMAN, C. L. *et al.* Knowledge infrastructures in science: data, diversity, and digital libraries. *International Journal on Digital Libraries*, v. 16, n. 3–4, p. 207–227, 2015. Disponível em: <http://link.springer.com/10.1007/s00799-015-0157-z>. Acesso em: 19 fev. 2018.

BORGMAN, C. L. *et al.* Data management in the long tail: science, software, and service. *International Journal of Digital Curation*, v. 11, n. 1, p. 128–149, 2016. Disponível em: <http://www.ijdc.net/index.php/ijdc/article/view/428>. Acesso em: 13 fev. 2018.

BORGMAN, C. L.; WALLIS, J. C.; ENYEDY, N. Little science confronts the data deluge: habitat ecology, embedded sensor networks, and digital libraries. *International Journal on Digital Libraries*, v. 7, n. 1–2, p. 17–30, 2007. Disponível em: <http://link.springer.com/10.1007/s00799-007-0022-9>. Acesso em: 25 ago. 2017.

BRANDT, D. S. Librarians as partners in e-research: Purdue University Libraries promote collaboration. *College & Research Libraries News*, v. 68, n. 6, p. 365–368, 2007. Disponível em: <http://crln.acrl.org/index.php/crlnews/article/view/7818/7818>. Acesso em: 6 jun. 2017.

BROWN, E. "I know what you researched last summer": how academic librarians are supporting researchers in the management of data curation. *The New Zealand Library & Information Management Journal*, v. 52, n. 1, p. 55–69, 2010.

BROWN, R. A.; WOLSKI, M.; RICHARDSON, J. Developing new skills for research support librarians. *The Australian Library Journal*, v. 64, n.3, p. 224–234, 2015. Disponível em: <http://www.tandfonline.com/doi/full/10.1080/00049670.2015.1041215>. Acesso em: 30 jul. 2018.

CALLAGHAN, S. *et al.* Making data a first class scientific output: data citation and publication by NERC's Environmental Data Centres. *International Journal of Digital Curation*, v. 7, n. 1, p. 107–113, 2012. Disponível em: <http://ijdc.net/index.php/ijdc/article/view/208>. Acesso em: 6 fev. 2018.

CANDELA, L. *et al.* Data journals: a survey. *Journal of the Association for Information*

Science and Technology, v. 66, n. 9, p. 1747–1762, 2015. Disponível em: <http://onlinelibrary.wiley.com/wol1/doi/10.1002/asi.23358/full>. Acesso em: 2 out. 2017.

CARLSON, J. The use of life cycle models in developing and supporting data services. In: RAY, J. M. (org.). *Research Data Management: practical strategies for information professionals*. West Lafayette: Purdue University Press, 2014. p. 63–86.

CARLSON, S. Lost in a sea of science data. *The Chronicle of Higher Education*, v. 52, n. 42, p. 35–36, 2006. Disponível em: <http://search.proquest.com.ezproxy.hioa.no/docview/214672107/1350FAF720E40E1CD2/1?accountid=26439>. Acesso em: 22 ago. 2017.

CASACUBERTA, D.; VALLVERDÚ, J. E-Science and the data deluge. *Philosophical Psychology*, v. 27, n. 1, p. 126–140, 2014. Disponível em: <http://www.tandfonline.com/doi/abs/10.1080/09515089.2013.827961>. Acesso em: 6 jun. 2017.

CESAR JUNIOR, R. M. Apresentação à edição brasileira. In: HEY, T.; STEWART, T.; TOLLE, K. (org.). *O quarto paradigma: descobertas científicas na era da eScience* Quarto paradigma: São Paulo: Oficina de Textos, 2011. p. 7–8.

CHIWARE, E. R. T.; BECKER, D. A. Research Data Management Services in Southern Africa: a readiness survey of academic and research libraries. *African Journal of Library, Archives and Information Science*, v. 28, n. 1, p. 1–16, 2018. Disponível em: <http://search-ebscohost-com.ez27.periodicos.capes.gov.br/login.aspx?direct=true&db=lih&AN=130261946&lang=pt-br&site=ehost-live&authtype=ip,cookie,uid>. Acesso em: 30 jul. 2018.

CHIWARE, E. R. T.; MATHE, Z. Academic libraries' role in Research Data Management Services: a South African perspective. *South African Journal of Libraries and Information Science*, v. 81, n. 2, p. 1–10, 2016. Disponível em: <http://sajlis.journals.ac.za/pub/article/view/1563>. Acesso em: 14 fev. 2019.

CLAIBOURN, M. P. Bigger on the inside: building Research Data Services at the University of Virginia. *Insights the UKSG Journal*, v. 28, n. 2, p. 100–106, 2015. Disponível em: <http://insights.uksg.org/articles/10.1629/uksg.239/>. Acesso em: 30 jul. 2018.

CLEMENTS, A. Research information meets research data management ... in the library? *Insights: the UKSG journal*, v. 26, n. 3, p. 298–304, 2013. Disponível em: <http://insights.uksg.org/articles/10.1629/2048-7754.99>. Acesso em: 30 jul. 2018.

CLOCKSS. *The CLOCKSS Archive: a trusted community-governed archive*. [S.l.], 2017. Disponível em: <https://www.clockss.org/clockss/Home>. Acesso em: 5 out. 2017.

CLUBB, J. M. et al. Sharing research data in the Social Sciences. In: FIENBERG, S. E.; MARTIN, M. E.; STRAF, M. L. (org.). *Sharing research data*. Washington, D.C.: National Academies Press, 1985. p. 39–80. Disponível em: <http://www.nap.edu/catalog/2033/sharing-research-data>. Acesso em: 25 set. 2017.

COATES, H. Building data services from the ground up: strategies and resources. *Journal of eScience Librarianship*, v. 3, n. 1, p. 52–59, 2014. Disponível em: <http://escholarship.umassmed.edu/jeslib/vol3/iss1/5>. Acesso em: 14 fev. 2019.

COMMITTEE ON DATA FOR SCIENCE AND TECHNOLOGY (CODATA). *UNISIST: study on the problems of accessibility and dissemination of data for science and technology*. [S.l.]: UNESCO, 1974. Disponível em: <http://unesdoc.unesco.org/images/0000/000095/009536eb.pdf>. Acesso em: 8 set. 2017.

CORRALL, S. Roles and responsibilities- libraries, librarians, and data. In: PRYOR, G. (org.). *Managing Research Data*. London: Facet Publishing, 2012. p. 105–133. Disponível em: http://d-scholarship.pitt.edu/25158/1/Corrall_%25282012%2529_RDM_chapter.pdf. Acesso em: 24 jul. 2017.

CORRALL, S.; KENNAN, M. A.; AFZAL, W. Bibliometrics and research data management services: emerging trends in library support for research. *Library Trends*, v. 61, n. 3, p. 636–674, 2013. Disponível em: <http://muse.jhu.edu/article/508619>. Acesso em: 6 jun. 2017.

CORRÊA, F. C. O papel dos bibliotecários na gestão de dados científicos. *Revista Digital de Biblioteconomia e Ciência da Informação*, v. 14, n. 3, p. 387–406, 2016. Disponível em: <http://dx.doi.org/10.20396/rdbci.v14i3.8646333>. Acesso em: 14 out. 2017.

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

CORTI, L. *et al.* The research data lifecycle. *Managing and sharing research data: a guide to good practice*. Los Angeles: Sage, 2014. p. 17–23.

COSTELLO, M. J. Motivating online publication of data. *BioScience*, v. 59, n. 5, p. 418–427, 2009. Disponível em: <https://academic.oup.com/bioscience/article-lookup/doi/10.1525/bio.2009.59.5.9>. Acesso em: 2 out. 2017.

COX, A. M. *et al.* Developments in research data management in academic libraries: towards an understanding of research data service maturity. *Journal of the Association for Information Science and Technology*, v. 68, n. 9, p. 2182–2200, 2017. Disponível em: <http://doi.wiley.com/10.1002/asi.23781>. Acesso em: 22 ago. 2017.

COX, A. M.; PINFIELD, S. Research data management and libraries: current activities and future priorities. *Journal of Librarianship and Information Science*, v. 46, n. 4, p. 299–316, 2014. Disponível em: <http://lis.sagepub.com/cgi/doi/10.1177/0961000613492542>. Acesso em: 24 ago. 2017.

COX, A. M.; VERBAAN, E.; SEM, B. Upskilling liaison librarians for research data management. *Ariadne*, n.70, 2012. Disponível em: <http://www.ariadne.ac.uk/issue/70/cox-et-al/>. Acesso em 28 ago. 2018.

CRAGIN, M. H. *et al.* Data sharing, small science and institutional repositories. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, v. 368, n. 1926, p. 4023–4038, 2010. Disponível em: <http://rsta.royalsocietypublishing.org/cgi/doi/10.1098/rsta.2010.0165>. Acesso em: 22 set. 2017.

CREAMER, A. *et al.* Assessment of health sciences and science and technology librarian e-science educational needs to develop an e-science web portal for librarians. *Journal of the Medical Library Association: JMLA*, v. 99, n. 2, p. 153–156, 2011. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3066590/>. Acesso em: 27 ago. 2018.

CREAMER, A. *et al.* An assessment of needed competencies to promote the data curation and management librarianship of Health Sciences and Science and Technology librarians in New England. *Journal of eScience Librarianship*, v. 1, n. 1, p. 18–26, 2012. Disponível em: <https://escholarship.umassmed.edu/jeslib/vol1/iss1/4/>. Acesso em: 27 ago. 2018.

CRESWELL, J. W. *Projeto de pesquisa: método qualitativo, quantitativo e misto*. Tradução Magda Lopes. 3. ed. Porto Alegre: Artmed, 2010.

CURTY, R. G. *Beyond data thrifting: an investigation of factors influencing research data re-use in the social sciences*. 2015. 295f. Tese (Doctor of Philosophy in Information Science

and Technology) - School of Information Studies, Syracuse University, Syracuse, 2015. Disponível em: <http://www.ijdc.net/index.php/ijdc/article/view/401>. Acesso em: 28 set. 2017.

CURTY, R. G. As diferentes dimensões do reuso de dados científicos. *Tendências da Pesquisa Brasileira em Ciência da Informação*, v. 9, n. 2, p. 1–28, 2016. Disponível em: <http://www.ufpb.br/evento/lti/ocs/index.php/enancib2016/enancib2016/paper/view/4056>. Acesso em: 28 set. 2017.

DATA CITATION SYNTHESIS GROUP. *Joint Declaration of Data Citation Principles*. San Diego, CA: FORCE11, 2014. Disponível em: <https://www.force11.org/datacitationprinciples>. Acesso em: 23 mar. 2018.

DATA DOCUMENTATION INITIATIVE (DDI). *Why Use DDI*. [S.l.], 2018. Disponível em: <http://www.ddialliance.org/training/why-use-ddi>. Acesso em: 19 jul. 2018.

DATA OBSERVATION NETWORK FOR EARTH (DataOne). *Data Life Cycle*. [S.l.], 2018. Disponível em: <<https://www.dataone.org/data-life-cycle>>. Acesso em: 19 jul. 2018.

DATA CITE. *Welcome to DataCite*. Disponível em: <<https://datacite.org/index.html>>. Acesso em: 13 fev. 2019.

DAUREEN, N. E-Science , cyberinfrastructure , and modern dance. *Information Outlook*, v. 12, n. 5, p. 42–44, 2008. Disponível em: https://www.sla.org/IO/2008/May/InfoOutlook_May08_Final.pdf. Acesso em: 5 jun. 2017.

DAVENHALL, C. *DCC Digital Curation Reference Manual: instalment on scientific metadata*. Edinburgh: DCC, 2011. Disponível em: <http://www.dcc.ac.uk/resources/curation-reference-manual/scientific-metadata>. Acesso em: 14 set. 2017.

DIGITAL CURATION CENTER (DCC). *What is digital curation*. Edinburgh, 2017. Disponível em: <http://www.dcc.ac.uk/digital-curation/what-digital-curation>. Acesso em: 14 set. 2017.

DIGITAL CURATION CENTRE (DCC). *DCC Curation Lifecycle Model*. Edinburgh, 2018a. Disponível em: <http://www.dcc.ac.uk/resources/curation-lifecycle-model>. Acesso em: 22 jul. 2018.

DIGITAL CURATION CENTER (DCC). *Disciplinary Metadata*. Edinburgh, 2018b. Disponível em: <http://www.dcc.ac.uk/resources/metadata-standards>. Acesso em: 11 maio 2018.

DIGITAL CURATION CENTRE (DCC). *Lifecycle Model FAQ*. Edinburgh, 2018c. Disponível em: <http://www.dcc.ac.uk/resources/curation-lifecycle-model/lifecycle-model-faqs>. Acesso em: 24 jul. 2018.

DUARTE, R. Entrevistas em pesquisa qualitativa. *Educar em Revista*, n.24, p. 213-225, 2004. Disponível em: <http://www.scielo.br/pdf/er/n24/n24a11.pdf>. Acesso em: 13 fev. 2019.

EDWARDS, P. N. *et al.* Science friction: data, metadata, and collaboration. *Social Studies of Science*, v. 41, n. 5, p. 667–690, 2011. Disponível em: <http://journals.sagepub.com/doi/10.1177/0306312711413314>. Acesso em: 19 fev. 2018.

E-INFRASTRUCTURE REFLECTION GROUP (E-IRG). *Long Tail of Data: e-IRG Task Force Report*. The Hague, NL: E-IRG, 2016. Disponível em: <http://e-irg.eu/documents/10920/238968/LongTailOfData2016.pdf>. Acesso em: 13 fev. 2018.

ELAG, M. M. *et al.* Identification and characterization of information-networks in long-tail data collections. *Environmental Modelling & Software*, v. 94, p. 100–111, 2017. Disponível em: <http://dx.doi.org/10.1016/j.envsoft.2017.03.032>. Acesso em: 13 fev. 2018.

ELECTRONIC LAB NOTEBOOK. In: WIKIPEDIA. *Wikipedia: the free encyclopedia*. [S.l.]: Wikimedia Foundation, 2017. Disponível em: https://en.wikipedia.org/wiki/Electronic_lab_notebook. Acesso em: 5 out. 2017.

ELO, S.; KYNGÄS, H. The qualitative content analysis process. *Journal of Advanced Nursing*, v. 62, n. 1, p. 107–115, 2008. Disponível em: <http://doi.wiley.com/10.1111/j.1365-2648.2007.04569.x>. Acesso em: 7 jan. 2019.

FAUNDEEN, J. L. *et al.* *The United States Geological Survey Science Data Lifecycle Model: U.S. Geological Survey Open-File Report 2013–1265*. [S.l.: s.n.], 2013. Disponível em: <http://dx.doi.org/10.3133/ofr20131265>. Acesso em: 23 jul. 2018.

FEARON JR, D. *et al.* *SPEC Kit 334: research data management services*. Washington, DC: ARL, 2013. Disponível em: <http://dx.doi.org/10.1016/j.acalib.2015.11.005>. Acesso em: 2 set. 2017.

FEDERATION OF EARTH SCIENCE INFORMATION PARTNERS (ESIP). *Data Citation Guidelines for Data Providers and Archives*. [S.l.]: ESIP Commons, 2012. Disponível em: <http://commons.esipfed.org/node/308>. Acesso em: 23 mar. 2018.

FERGUSON, A. R. *et al.* Big data from small data: data-sharing in the “long tail” of neuroscience. *Nature Neuroscience*, v. 17, n. 11, p. 1442–1447, 2014. Disponível em: [http://dx.doi.org/10.1038/nn.3838fv=\[5tr](http://dx.doi.org/10.1038/nn.3838fv=[5tr). Acesso em: 6 mar. 2018.

FIENBERG, S. E.; MARTIN, M. E.; STRAF, M. L. (org.). *Sharing research data*. Washington, D.C.: National Academy Press, 1985. Disponível em: <http://www.nap.edu/catalog/2033/sharing-research-data>. Acesso em: 25 set. 2017.

FLORES, J. R. *et al.* Libraries and the research data management landscape. In: MACLACHLAN, J. C.; WARAKSA, E. A.; WILLIFORD, C. (org.). *The Process of Discovery: The CLIR Postdoctoral Fellowship Program and the Future of the Academy*. [S.l.]: Council on Library and Information Resources, 2015. v. 167. p. 82–102. Disponível em: <http://www.clir.org/pubs/reports/pub167/RDM.pdf>. Acesso em: 18 ago. 2017.

FONTANELLA, B. J. B. *et al.* Amostragem em pesquisas qualitativas: proposta de procedimentos para constatar saturação teórica. *Cadernos de Saúde Pública*, v. 27, n. 2, p. 388–394, 2011. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0102-311X2011000200020&lng=pt&tlng=pt. Acesso em: 18 fev. 2018.

FOX, P.; HENDLER, J. eScience semântica: o significado codificado na próxima geração de ciência digitalmente aprimorada. In: HEY, T.; STEWART, T.; TOLLE, K. (org.). *O quarto paradigma: descobertas científicas na era da eScience*. São Paulo: Oficina de Textos, 2011. p. 159–163.

FOX, R. Digital libraries: the systems analysis perspective. *OCLC Systems & Services: International digital library perspectives*, v. 28, n. 4, p. 170–175, 2012. Disponível em: <http://www.emeraldinsight.com/doi/10.1108/10650751211279102>. Acesso em: 30 jul. 2018.

FREDERICK, D. E. Data, open science and libraries: the data deluge column. *Library Hi Tech News*, v. 33, n. 8, p. 11–16, 2016. Disponível em: <http://www.emeraldinsight.com/doi/10.1108/LHTN-09-2016-0040>. Acesso em: 31 ago. 2017.

FRIEDLANDER, A.; ADLER, P. *To stand the test of time: long-term stewardship of digital data sets in Science and Engineering*. Washington, D.C.: ARL, 2006. Disponível em: <http://www.arl.org/storage/documents/publications/digital-data-report-2006.pdf>. Acesso em: 20 abr. 2018.

FUNDAÇÃO DE AMPARO À PESQUISA DO ESTADO DE MINAS GERAIS (FAPEMIG). *Diretrizes e informações gerais*. Belo Horizonte: FAPEMIG, [2019]. Disponível em: <http://www.fapemig.br/visualizacao-de-normas-gerais/ler/370/diretrizes-e-informacoes-gerais>. Acesso em: 13 maio 2019.

FUNDAÇÃO DE AMPARO À PESQUISA DO ESTADO DE SÃO PAULO (FAPESP). *Plano de Gestão de Dados - FAPESP*. São Paulo: FAPESP, [2019]. Disponível em: <http://www.fapesp.br/gestaodedados/>. Acesso em: 13 maio 2019.

GABRIDGE, T. The last mile: liaison roles in curating Science and Engineering research data. *Research Library Issues: a bimonthly report from ARL, CNI, and SPARC*, n. 265, p. 15–21, 2009. Disponível em: <http://0-search.ebscohost.com.umaclib3.umac.mo/login.aspx?direct=true&db=lih&AN=44682079&site=eds-live>. Acesso em: 24 ago. 2017.

GARRITANO, J. R.; CARLSON, J. R. A subject librarian's guide to collaborating on e-science projects. *Issues in Science and Technology Librarianship*, v. 57, p. 1–15, 2009. Disponível em: <http://www.istl.org/09-spring/refereed2.html>. Acesso em: 27 ago. 2018.

GERACI, D.; HUMPHREY, C.; JACOBS, J. *Data basics: an introductory text*. Ann Arbor: Inter-university Consortium for Political and Social Research (ICPSR), 2012. Disponível em: <https://doi.org/10.7939/R3251FK8F>. Acesso em: 14 set. 2018.

GERMAN FEDERATION FOR BIOLOGICAL DATA (GFBIO). *Data Life cycle*. Bremen: GFBIO, [2018]. Disponível em: <https://www.gfbio.org/pt/training/materials/data-lifecycle>. Acesso em: 25 jul. 2018.

GIARLO, M. J. Academic libraries as data quality hubs. *Journal of Librarianship and Scholarly Communication*, v. 1, n. 3, p. e1059 1-10, 2013. Disponível em: <http://jlscc-pub.org/articles/10.7710/2162-3309.1059>. Acesso em: 14 fev. 2019.

GIBBS, G. *Análise de dados qualitativos*. Porto Alegre: Artmed, 2009.

GIL, A. C. *Métodos e técnicas de pesquisa social*. 4. ed. São Paulo: Atlas, 1994.

GOBEN, A.; NELSON, M. R. S. The Data Engagement Opportunities Scaffold: Development and Implementation. *Journal of eScience Librarianship*, v. 7, n. 2, p. e1128 1-7, 2018. Disponível em: <https://escholarship.umassmed.edu/cgi/viewcontent.cgi?article=1128&context=jeslib>. Acesso em: 27 ago. 2018.

GOLD, A. Cyberinfrastructure, data, and libraries, Part 1: a cyberinfrastructure primer for librarians. *D-Lib Magazine*, v. 13, n. 9/10, online, 2007a. Disponível em: <http://www.dlib.org/dlib/september07/gold/09gold-pt1.html>. Acesso em: 27 ago. 2017.

GOLD, A. Cyberinfrastructure, data, and libraries, Part 2: libraries and the data challenge: roles and actions for libraries. *D-Lib Magazine*, v. 13, n. 9/10, online, 2007b. Disponível em: <http://www.dlib.org/dlib/september07/gold/09gold-pt2.html>. Acesso em: 27 ago. 2017.

GOLDMAN, J.; KAFEL, D.; MARTIN, E. Assessment of data management services at New England Region Resource Libraries. *Journal of eScience Librarianship*, v. 4, n. 1, p. e1068 1-17, 2015. Disponível em: <http://escholarship.umassmed.edu/jeslib/vol4/iss1/4/>. Acesso em: 22 ago. 2018.

GORE, S. A. e-Science and data management resources on the Web. *Medical reference services quarterly*, v. 30, n. 2, p. 167–77, 2011. Disponível em: <http://www.scopus.com/inward/record.url?eid=2-s2.0-79957541775&partnerID=tZotx3y1>.

GRAY, J. *et al.* Online scientific data curation, publication, and archiving. [S.l.]: Virtual Observatories, 2002. p. 103–107. Disponível em: <http://arxiv.org/abs/cs/0208012>. Acesso em: 20 mar. 2018.

GUPTILL, S. C. Metadata and data catalogues. *In*: LONGLEY, P. A. *et al.* (org.). *Geographical Information Systems: principles and technical issues*. New York: John Wiley & Sons, 1999. v. 2. p. 677–692.

HALL, W.; DE ROURE, D.; SHADBOLT, N. The evolution of the Web and implications for eResearch. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, v. 367, n. 1890, p. 991–1001, 2009. Disponível em: <http://rsta.royalsocietypublishing.org/cgi/doi/10.1098/rsta.2008.0252>. Acesso em: 27 ago. 2017.

HARRIS, S. Distinctive services in academic librarianship. *New Library World*, v. 117, n. 9/10, p. 596–625, 2016. Disponível em: <http://www.emeraldinsight.com/doi/10.1108/NLW-05-2016-0036>. Acesso em: 14 fev. 2019.

HAYES, S. eScience. What Is This? And, what does it have to do with hospital librarians. *Journal of Hospital Librarianship*, v. 14, n. 4, p. 402–408, 2014. Disponível em: [10.1080/15323269.2014.951423%255Cnhttp://libproxy.sdsu.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=aph&AN=99237128&site=ehost-live](http://libproxy.sdsu.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=aph&AN=99237128&site=ehost-live). Acesso em: 13 fev. 2018.

HEIDORN, P. B. Shedding light on the dark data in the long tail of science. *Library Trends*, v. 57, n. 2, p. 280–299, 2008. Disponível em: http://muse.jhu.edu/content/crossref/journals/library_trends/v057/57.2.heidorn.html. Acesso em: 13 fev. 2018.

HEIDORN, P. B. The emerging role of libraries in data curation and e-science. *Journal of Library Administration*, v. 51, n. 7–8, p. 662–672, 2011. Disponível em: <http://www.tandfonline.com/doi/abs/10.1080/01930826.2011.601269>. Acesso em: 27 ago. 2017.

HENDERSON, M. E.; KNOTT, T. L. Starting a research data management program based in a university Library. *Medical Reference Services Quarterly*, v. 34, n. 1, p. 47–59, 2015. Disponível em: <http://www.tandfonline.com/doi/abs/10.1080/02763869.2015.986783>. Acesso em: 30 jul. 2018.

HENTY, M. Dreaming of data: the library's role in supporting e-Research and data management. Alice Springs, NT Australia: ALIA, 2008. p. 1–10. Disponível em: http://apsr.anu.edu.au/presentations/henty_alia_08.pdf. Acesso em: 30 ago. 2017.

HEY, T.; HEY, J. e-Science and its implications for the library community. *Library Hi Tech*, v. 24, n. 4, p. 515–528, 2006. Disponível em: <http://www.emeraldinsight.com/doi/10.1108/07378830610715383>. Acesso em: 6 jun. 2017.

HEY, T.; TANSLEY, S.; TOLLE, K. (org.). *O Quarto Paradigma: descobertas científicas na era da eScience*. São Paulo: Oficina de Textos, 2011.

HEY, T.; TREFETHEN, A. E. The UK e-Science Core Programme and the Grid. *Future Generation Computer Systems*, v. 18, n. 8, p. 1017–1031, out. 2002. Disponível em: <http://linkinghub.elsevier.com/retrieve/pii/S0167739X02000821>. Acesso em: 6 set. 2017.

HEY, T.; TREFETHEN, A. E. The Data Deluge: an e-Science Perspective. *In*: BERMAN, F.; FOX, G.; HEY, T. (org.). *Grid Computing: making the global infrastructure a reality*. Chichester, UK: John Wiley & Sons, 2003. p. 809–824. Disponível em:

<http://doi.wiley.com/10.1002/0470867167.ch36>. Acesso em: 22 ago. 2017.

HEY, T.; TREFETHEN, A. E. Cyberinfrastructure for e-Science. *Science*, v. 308, n. 5723, p. 817–821, 2005. Disponível em: <http://www.sciencemag.org/cgi/doi/10.1126/science.1110410>. Acesso em: 5 jun. 2017.

HIGGINS, S. *Using Metadata Standards. DCC Briefing Papers*. Edinburgh: DCC, 2007a. Disponível em: <http://www.dcc.ac.uk/resources/briefing-papers/standards-watch-papers/using-metadata-standards>. Acesso em: 10 maio 2018.

HIGGINS, S. *What are Metadata Standards. DCC Briefing Papers*. Edinburgh: DCC, 2007b. Disponível em: <http://www.dcc.ac.uk/resources/briefing-papers/standards-watch-papers/what-are-metadata-standards>. Acesso em: 10 maio 2018.

HIGGINS, S. The DCC Curation Lifecycle Model. *The International Journal of Digital Curation*, v. 3, n. 1, p. 134–140, 2008. Disponível em: <http://www.ijdc.net/index.php/ijdc/article/view/69/48>. Acesso em: 17 set. 2017.

HINE, C. (org.). *New Infrastructures for Knowledge Production: Understanding E-science*. Hershey: Information Science Publishing, 2006. Disponível em: <http://books.google.com/books?id=exyQGONs4VgC&pgis=1>. Acesso em: 17 set. 2017.

HIOM, D. et al. Research data management at the University of Bristol. *Program: electronic library and information systems*, v. 49, n. 4, p. 475–493, 2015. Disponível em: <http://www.emeraldinsight.com/doi/10.1108/PROG-02-2015-0019>. Acesso em: 29 maio 2018.

HIOM, D. et al. Introducing safe access to sensitive data at the University of Bristol. *International Journal of Digital Curation*, v. 12, n. 2, p. 26–36, 2017. Disponível em: <http://www.ijdc.net/article/view/506>. Acesso em 01 ago. 2018.

HSIEH, H.; SHANNON, S. E. Three approaches to qualitative content analysis. *Qualitative Health Research*, v. 15, n. 9, p. 1277–1288, 2005. Disponível em: <http://journals.sagepub.com/doi/10.1177/1049732305276687>. Acesso em: 7 jan. 2019.

HUMPHREY, C. *e-Science and the Life Cycle of Research*. Edmonton: [s.n.], 2006. Disponível em: <https://era.library.ualberta.ca/items/3334684b-fa6a-4c9d-a74b-559fec42f9f>. Acesso em: 01 ago. 2018.

INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA (IBICT). *Sobre Repositórios Digitais*. IBICT: [Brasília], 2018. Disponível em: <http://www.ibict.br/informacao-para-a-pesquisa/repositorios-digitais>. Acesso em: 4 maio 2018.

INTERNATIONAL FEDERATION OF LIBRARY ASSOCIATIONS (IFLA). *Best Practice for National Bibliographic Agencies in a Digital Age*. IFLA: [S.l.], 2019. Disponível em: <https://www.ifla.org/node/7858>. Acesso em: 13 fev. 2019.

INTERNATIONAL DOI FOUNDATION (IDF). *DOI*. IDF: [S.l.], 2018. Disponível em: <https://www.doi.org>. Acesso em: 13 fev. 2019.

JAHNKE, L.; ASHER, A.; KERALIS, S. D. C. *The problem of data*. Washington, DC: Council on Library and Information Resources (CLIR), 2012. Disponível em: <https://www.clir.org/pubs/reports/pub154/pub154.pdf>. Acesso em: 31 ago. 2017.

JANKOWSKI, N. W. Exploring e-Science: an introduction. *Journal of Computer-Mediated Communication*, v. 12, n. 2, p. 549–562, 2007. Disponível em: <http://doi.wiley.com/10.1111/j.1083-6101.2007.00337.x>. Acesso em: 6 jun. 2017.

JOHNSON, L. M.; BUTLER, J. T.; JOHNSTON, L. R. Developing e-Science and research services and support at the University of Minnesota Health Sciences Libraries. *Journal of Library Administration*, v. 52, n. 8, p. 754–769, 2012. Disponível em: <http://www.tandfonline.com/doi/abs/10.1080/01930826.2012.751291>. Acesso em: 27 ago. 2018.

JOINT, N. Data preservation, the new science and the practitioner librarian. *Library Review*, v. 56, n. 6, p. 451–455, 2007. Disponível em: <http://www.emeraldinsight.com/doi/10.1108/00242530710760337>. Acesso em: 16 set. 2017.

JONES, S.; PRYOR, G.; WHYTE, A. *How to Develop RDM Services: a guide for HEIs*. Edinburgh: DCC, 2013. Disponível em: <http://www.dcc.ac.uk/resources/how-guides/how-develop-rdm-services>. Acesso em: 3 jun. 2018. (DCC How-to Guides).

KENNAN, M. A.; CORRALL, S.; AFZAL, W. “Making space” in practice and education: research support services in academic libraries. *Library Management*, v. 35, n. 8/9, p. 666–683, 2014. Disponível em: <http://www.emeraldinsight.com/doi/10.1108/LM-03-2014-0037>. Acesso em: 18 ago. 2017.

KERBY, E. E. Research data services in veterinary medicine libraries. *Journal of the Medical Library Association : JMLA*, v. 104, n. 4, p. 305–308, 2016. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5079493/%0Ahttps://www.scopus.com/inward/record.uri?eid=2-s2.0-84994763019&partnerID=40&md5=2ecc19787175ad12dd0db78d259ec90d>. Acesso em: 30 jul. 2018.

KAFEL, D.; CREAMER, A.; MARTIN, E. Building the New England Collaborative Data Management Curriculum. *Journal of eScience Librarianship*, v. 3, n. 1, p. 60–66, 2014. Disponível em: <http://escholarship.umassmed.edu/jeslib/vol3/iss1/7>. Acesso em: 01 ago. 2018.

KIM, J. Data sharing and its implications for academic libraries. *New Library World*, 2013. Disponível em: <https://www.emerald.com/insight/content/doi/10.1108/NLW-06-2013-0051/full/pdf?title=data-sharing-and-its-implications-for-academic-libraries>. Acesso em: 30 jul. 2018.

KITCHIN, R.; LAURIAULT, T. P. Small data in the era of big data. *GeoJournal*, v. 80, n. 4, p. 463–475, 2015. Disponível em: <http://dx.doi.org/10.1007/s10708-014-9601-7>. Acesso em: 6 mar. 2018.

KNIGHT, G. Building a research data management service for the London school of hygiene & tropical medicine. *Program: electronic library and information systems*, v. 49, n. 4, p. 424–439, 2015. Disponível em: <https://www.emeraldinsight.com/doi/full/10.1108/PROG-01-2015-0011>. Acesso em: 30 jul. 2018.

KNUTH, S. L. et al. The Center for Research Data and Digital Scholarship at the University of Colorado-Boulder. *Bulletin of the Association for Information Science and Technology*, v. 43, n. 2, p. 46–48, 2017. Disponível em: <http://doi.wiley.com/10.1002/bul2.2017.1720430215>. Acesso em: 30 jul. 2018.

KNUTH, S. L.; JOHNSON, A. M.; HAUSER, T. Research Data Services at the University of Colorado Boulder. *Bulletin of the Association for Information Science and Technology*, v. 41, n. 6, p. 35–38, ago. 2015. Disponível em: <https://www.asist.org/publications/bulletin/aug-2015/research-data-services/>. Acesso em: 30 jul. 2018.

KOLTAY, T. Data literacy: in search of a name and identity. *Journal of Documentation*, v. 71, n. 2, p. 401–415, 2015. Disponível em: <http://www.emeraldinsight.com/doi/10.1108/JD-02->

2014-0026. Acesso em: 30 jul. 2018.

KOLTAY, T. Are you ready? Tasks and roles for academic libraries in supporting Research 2.0. *New Library World*, v. 117, n. 1/2, p. 94–104, 2016a. Disponível em: <http://www.emeraldinsight.com/doi/10.1108/NLW-09-2015-0062>. Acesso em: 30 jul. 2018.

KOLTAY, T. Digital Research Data: Where are we Now. In: BAKER, D.; EVANS, W. (org.). *Digital Information Strategies: from applications and content to libraries and people*. Waltham, MA: Chandos Publishing, 2016b. p. 71–84. Disponível em: <https://doi.org/10.1016/b978-0-08-100251-3.00005-6>. Acesso em: 15 ago. 2017.

KOLTAY, T. Research 2.0 and Research Data Services in academic and research libraries: priority issues. *Library Management*, v. 38, n. 6/7, p. 345–353, 15 ago. 2017. Disponível em: <http://www.emeraldinsight.com/doi/10.1108/LM-11-2016-0082>. Acesso em: 18 ago. 2017.

KOLTAY, T.; ŠPIRANEC, S.; KARVALICS, L. Z. The shift of information literacy towards Research 2.0. *Journal of Academic Librarianship*, v. 41, n. 1, p. 87–93, 2015. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0099133314001979>. Acesso em: 18 ago. 2017.

KOWALCZYK, S. T. Modelling the Digital Research Data Lifecycle. *International Journal of Digital Curation*, v. 12, n. 2, p. 331–361, 2017. Disponível em: <https://doi.org/10.2218/ijdc.v12i2.429>. Acesso em: 13 fev. 2018.

KOWALCZYK, S. T. *Digital Curation for Libraries and Archives*. Santa Barbara, California: Libraries Unlimited, 2018.

KRAFT, A. RADAR - A repository for long tail data. Hanover, DE: IATUL, 2015. Disponível em: <https://docs.lib.purdue.edu/cgi/viewcontent.cgi?article=2099&context=iatul>. Acesso em: 13 fev. 2018.

KRATZ, J. E.; STRASSER, C. Data publication consensus and controversies. *F1000Research*, n. 3, p. 1–21, 23 abr. 2014. Disponível em: <http://f1000research.com/articles/3-94/v1>. Acesso em: 28 set. 2017.

KRATZ, J. E.; STRASSER, C. Researcher perspectives on publication and peer review of data. *PLOS ONE*, v. 10, n. 2, p. e0117619 1–21, 2015. Disponível em: <http://dx.plos.org/10.1371/journal.pone.0117619>. Acesso em: 18 ago. 2017.

LATHAM, B. Research Data Management: defining roles, prioritizing services, and enumerating challenges. *The Journal of Academic Librarianship*, v. 43, n. 3, p. 263–265, 2017. Disponível em: <http://linkinghub.elsevier.com/retrieve/pii/S0099133317301453>. Acesso em: 2 set. 2017.

LAVILLE, C.; DIONNE, J.; SIMAN, L. M. DE C. *A construção do saber: manual de metodologia da pesquisa em ciências humanas*. Porto Alegre: Artmed, 1999.

LAWRENCE, B. et al. Citation and peer review of data: moving towards formal data publication. *International Journal of Digital Curation*, v. 6, n. 2, p. 4–37, 2011. Disponível em: <http://www.ijdc.net/index.php/ijdc/article/view/181/265>. Acesso em: 28 set. 2017.

LEE, D. J.; STVILIA, B. Practices of research data curation in institutional repositories: a qualitative view from repository staff. *PLoS ONE*, v. 12, n. 3, p. e0173987 1–44, 2017. Disponível em: <http://dx.plos.org/10.1371/journal.pone.0173987>. Acesso em: 16 set. 2017.

LEITE, F. C. L. *Como gerenciar e ampliar a visibilidade da informação científica brasileira: repositórios institucionais de acesso aberto*. Brasília: IBICT, 2009. Disponível em: http://repositorio.unb.br/bitstream/10482/4841/1/LIVRO_ComoAmpliarGerenciar.pdf.

Acesso em: 26 maio 2018.

LEWIS, M. J. Libraries and the management of research data. In: MCKNIGHT, S. (org.). *Envisioning future academic library services: initiatives, ideas and challenges*. London: Facet Publishing, 2010. p. 145–168. Disponível em: <http://www.facetpublishing.co.uk/title.php?id=691-6>. Acesso em: 22 ago. 2017.

LIBARDONI, M. Fundamentos teóricos e visão estratégica da advocacy. *Revista Estudos Feministas*, v. 8, n. 2, p. 207–222, 2000. Disponível em: <https://periodicos.ufsc.br/index.php/ref/article/view/11936>. Acesso em: 13 fev. 2019.

LIU, X.; DING, N. Research data management in universities of central China. *The Electronic Library*, v. 34, n. 5, p. 808–822, 2016. Disponível em: <http://www.emeraldinsight.com/doi/10.1108/EL-04-2015-0063>. Acesso em: 14 fev. 2019.

LORD, P.; MACDONALD, A. *e-Science Data Curation*: [S.l.]: JISC, 2004. Disponível em: <http://www.cs.york.ac.uk/ftpdir/pub/leo/york-msc-2007/information/vsr-curation/science-dc-report.pdf%255Cnhttps://drive.google.com/file/d/0BwVuBK4FRW2zT2RYelhSTGFxODg/edit?usp=sharing>. Acesso em: 16 set. 2017.

LUCE, R. E. A new value equation challenge: the emergence of eResearch and roles for research libraries. Washington, D.C.: CLIR, 2008. p. 42–50. Disponível em: <https://www.clir.org/pubs/reports/pub142/luce.html>. Acesso em: 2 jul. 2017.

LYNCH, C. A. Institutional repositories: essential infrastructure for scholarship in the digital age. *ARL: A Bimonthly Report*, n. 226, 2003. Disponível em: <http://old.arl.org/resources/pubs/br/br226/br226ir.shtml>. Acesso em: 26 maio 2018.

LYNCH, C. A. Big data: how do your data grow? *Nature*, v. 455, n. 7209, p. 28–29, 2008. Disponível em: <http://www.nature.com/nature/journal/v455/n7209/full/455028a.html?foxtrotcallback=true>. Acesso em: 31 ago. 2017.

LYON, L. *Dealing with data*: roles, rights, responsibilities and relationships. Bath: UKOLN, 2007. Disponível em: https://purehost.bath.ac.uk/ws/portalfiles/portal/419529/dealing_with_data_report-final.pdf. Acesso em: 31 ago. 2017.

LYON, L. The informatics transform: re-engineering libraries for the data decade. *International Journal of Digital Curation*, v. 7, n. 1, p. 126–138, 2012. Disponível em: <http://www.ijdc.net/index.php/ijdc/article/view/210/279>. Acesso em: 23 ago. 2017.

MACDONALD, S.; MARTINEZ-URIBE, L. Data librarianship: a gap in the market. *Update*, v. 7, n. 6, p. 20–21, 2013. Disponível em: [https://www.era.lib.ed.ac.uk/bitstream/handle/1842/2499/Gap in the market.pdf?sequence=3&isAllowed=y](https://www.era.lib.ed.ac.uk/bitstream/handle/1842/2499/Gap%20in%20the%20market.pdf?sequence=3&isAllowed=y). Acesso em: 6 out. 2017.

MACMILLAN, D. Data sharing and discovery: what librarians need to know. *Journal of Academic Librarianship*, v. 40, n. 5, p. 541–549, 2014. Disponível em: <http://dx.doi.org/10.1016/j.acalib.2014.06.011>. Acesso em: 2 set. 2017.

MANYIKA, J. *et al.* *Big data: the next frontier for innovation, competition, and productivity*. [S.l.]: McKinsey Global Institute, 2011. Disponível em: <https://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/big-data-the-next-frontier-for-innovation>. Acesso em: 27 ago. 2017.

MÁRDERO ARELLANO, M. A. Preservação e curadoria de dados científicos. In: CONFERÊNCIA SOBRE TECNOLOGIA, CULTURA E MEMÓRIA, 2017, Recife.

Apresentações. Recife: UFPE, 2017. Disponível em: <http://www.liber.ufpe.br/ctcm2017/>. Acesso em: 4 out. 2017.

MARTIN, E. The role of librarians in data science: a call to action. *Journal of eScience Librarianship*, v. 4, n. 2, p. e1092 1-3, 2016. Disponível em: <http://escholarship.umassmed.edu/jeslib/vol4/iss2/7/>. Acesso em: 5 out. 2017.

MARTINEZ, L.; MACDONALD, S. Supporting local data users in the UK Academic Community. *Ariadne*, n. 44, 2005. Disponível em: <http://www.ariadne.ac.uk/issue44/martinez>. Acesso em: 16 out. 2017.

MATUSIAK, K. K.; SPOSITO, F. A. Types of research data management services: an international perspective. *Proceedings of the Association for Information Science and Technology*, v. 54, n. 1, p. 754–756, 2017. Disponível em: <http://doi.wiley.com/10.1002/pr2.2017.14505401144>. Acesso em: 18 abr. 2018.

MCLURE, M. *et al.* Data curation: a study of researcher practices and needs. *Portal: Libraries and the Academy*, v. 14, n. 2, p. 139–164, 2014. Disponível em: <https://dspace.library.colostate.edu/bitstream/handle/10217/81401/14.2.mclure.pdf;sequence=1>. Acesso em: 16 set. 2017.

MERZ, M. Embedding digital infrastructure in epistemic culture. In: HINE, C. (org.). *New Infrastructures for Knowledge Production: Understanding E-science*. Hershey: Information Science Publishing, 2006. p. 99–118.

MICHEL, M. H. *Metodologia e pesquisa científica em Ciências Sociais*. 2. ed. São Paulo: Atlas, 2009.

MINAYO, M. C. DE S. Amostragem e saturação em pesquisa qualitativa: consensos e controvérsias. *Revista Pesquisa Qualitativa*, v. 5, n. 7, p. 1–12, 2017. Disponível em: <http://rpq.revista.sepq.org.br/index.php/rpq/article/view/82/59>. Acesso em: 26 nov. 2017.

MONASTERSKY, R. Publishing frontiers: the library reboot. *Nature*, v. 495, n. 7442, p. 430–432, 2013. Disponível em: <http://www.nature.com/doi/10.1038/495430a>. Acesso em: 24 ago. 2017.

MULLINS, J. L. Bringing librarianship to e-Science. *College & Research Libraries*, v. 70, n. 3, p. 212–213, 2009. Disponível em: <http://0-search.ebscohost.com/libraries.colorado.edu/login.aspx?direct=true&db=ofm&AN=502977633&site=ehost-live&scope=site>. Acesso em: 24 ago. 2017.

NATIONAL LIBRARY OF MEDICINE. *Common Data Element (CDE) Resource Portal: Glossary*. NLM: Bethesda, 2013. Disponível em: <https://www.nlm.nih.gov/cde/glossary.html>. Acesso em: 11 maio 2018.

NATIONAL NETWORK OF LIBRARIES OF MEDICINE (NNLM). *Data Thesaurus*. NNLM: Bethesda, [2018]. Disponível em: <https://nnlm.gov/data/thesaurus>. Acesso em: 20 out. 2018.

NATIONAL RESEARCH COUNCIL (NRC). Importance and use of scientific and technical databases. In: NATIONAL RESEARCH COUNCIL (NRC). *A Question of Balance: Private Rights and the Public Interest in Scientific and Technical Databases*. Washington, DC: The National Academies Press, 1999. p. 14–39.

NATIONAL RESEARCH COUNCIL (NRC). Cyberinfrastructure. In: NATIONAL RESEARCH COUNCIL. *Transforming Combustion Research through Cyberinfrastructure*. Washington, D.C.: National Academies Press, 2011. p. 14–40. Disponível em: <http://www.nap.edu/catalog/13049>. Acesso em: 20 jun. 2017.

NATIONAL SCIENCE BOARD (NSB). *Long-lived digital data collections: enabling research and education in the 21st Century*. Arlington: NSF, 2005. Disponível em: <http://www.nsf.gov/pubs/2005/nsb0540/nsb0540.pdf>. Acesso em: 22 jun. 2017.

NATIONAL SCIENCE FOUNDATION (NSF). *Revolutionizing Science and Engineering through cyberinfrastructure*: report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure. Arlington: NSF, 2003. Disponível em: <http://www.nsf.gov/od/oci/reports/atkins.pdf>. Acesso em: 20 jun. 2017.

NATIONAL SCIENCE FOUNDATION (NSF). *Cyberinfrastructure Vision for 21st Century Discovery*. [S.l.]: NSF, 2007. Disponível em: <http://www.nsf.gov/pubs/2007/nsf0728/nsf0728.pdf>. Acesso em: 7 set. 2017.

NENTWICH, M. Conceptual framework: definitions and model. In: NENTWICH, M. *Cyberscience: research in the age of the internet*. Vienna: Austrian Academy of Sciences Press, 2003. p. 21–64.

NEWMAN, P.; CORKE, P. Data papers: peer reviewed publication of high quality data. *International Journal of Robotics Research*, v. 28, n. 5, p. 587, 2009. Disponível em: <http://journals.sagepub.com/doi/10.1177/0278364909104283>. Acesso em: 2 out. 2017.

NEWTON, M. P.; MILLER, C. C.; BRACKE, M. S. Librarian roles in institutional repository data set collecting: outcomes of a research library task force. *Collection Management*, v. 36, n. 1, p. 53–67, 2010. Disponível em: <http://www.tandfonline.com/doi/abs/10.1080/01462679.2011.530546>. Acesso em: 24 ago. 2017.

O'SHEA, D. C. Dark data. *Optical Engineering*, v. 47, n. 1, p. 010101, 2008. Disponível em: <https://www.spiedigitallibrary.org/journals/Optical-Engineering/volume-47/issue-01/010101/Dark-Data/10.1117/1.2838027.full?SSO=1>. Acesso em: 14 fev. 2018.

OLIVEIRA, D. M. T. Amostra qualitativa e recrutamento. In: PERDIGÃO, D. M.; HERLINGER, M.; WHITE, O. M. (org.). *Teoria e prática da pesquisa aplicada*. Rio de Janeiro: Elsevier, 2012. p. 120–127.

ONSRUD, H.; CAMPBELL, J. Big opportunities in access to “Small Science” data. *Data Science Journal*, v. 6, p. 58–66, 2007. Disponível em: <http://datascience.codata.org/articles/abstract/10.2481/dsj.6.OD58/>. Acesso em: 18 fev. 2018.

ORCID. *Open Researcher and Contributor ID*. ORCID: [S.l.], [2019]. Disponível em: <https://orcid.org/>. Acesso em: 13 fev. 2019.

OSSWALD, A. E-science and information services: a missing link in the context of digital libraries. *Online Information Review*, v. 32, n. 4, p. 516–523, 2008. Disponível em: <http://www.emeraldinsight.com/doi/10.1108/14684520810897395>. Acesso em: 20 set. 2017.

OU, S.; ZHOU, Y. Current status of scientific data curation research and practices in Mainland China. *LIBRES: Library and Information Science Research Electronic Journal*, v. 26, n. 1, p. 73–88, 2016. Disponível em: <http://www.libres-ejournal.info/2398/>. Acesso em: 16 set. 2017.

PALMER, C. L. et al. Data curation for the long tail of science: the case of environmental sciences. In: INTERNATIONAL DIGITAL CURATION CONFERENCE, 3., 2007, Washington, DC. *Anais [...]*. Edinburgh: DCC, 2007. p. 1–5. Disponível em: https://apps.lis.illinois.edu/wiki/download/attachments/32666/Palmer_DCC2007.pdf. Acesso em: 18 fev. 2018.

PARSONS, M. A.; FOX, P. A. Is Data publication the right metaphor? *Data Science Journal*, v. 12, p. 32-46, 2013. Disponível em: <http://datascience.codata.org/articles/abstract/10.2481/dsj.WDS-042/>. Acesso em: 22 mar. 2018.

PATERSON, M. *et al.* DART: a new missile in Australia's e-research strategy. *Online Information Review*, v. 31, n. 2, p. 116–134, 2007. Disponível em: <http://www.emeraldinsight.com.ezproxy.hio.no/journals.htm?issn=1468-4527&volume=31&issue=2&articleid=1602558&show=html>. Acesso em: 5 jun. 2017.

PERDIGÃO, D. M. Introdução à pesquisa quantitativa. In: PERDIGÃO, D. M.; HERLINGER, M.; WHITE, O. M. (org.). *Teoria e prática da pesquisa aplicada*. Rio de Janeiro: Elsevier, 2012. p. 109–117.

PETERS, C.; VAUGHN, P. Initiating data management instruction to graduate students at the University of Houston using the New England Collaborative Data Management Curriculum. *Journal of eScience Librarianship*, v. 3, n. 1, p. 86–99, 2014. Disponível em: <http://escholarship.umassmed.edu/jeslib/vol3/iss1/11>. Acesso em: 30 jul. 2018.

PFEIFFENBERGER, H.; CARLSON, D. "Earth System Science Data" (ESSD): peer reviewed journal for publication of data. *D-Lib Magazine*, v. 17, n. 1/2, p. 7–11, jan. 2011. Disponível em: <http://www.dlib.org/dlib/january11/pfeiffenberger/01pfeiffenberger.html>. Acesso em: 2 out. 2017.

PINFIELD, S.; COX, A. M.; SMITH, J. Research data management and libraries: relationships, activities, drivers and influences. *PLoS ONE*, v. 9, n. 12, p. e114734 1-28, 2014. Disponível em: <http://dx.plos.org/10.1371/journal.pone.0114734>. Acesso em: 8 maio 2018.

POOLE, A. H. How has your science data grown? Digital curation and the human factor: a critical literature review. *Archival Science*, v. 15, n. 2, p. 101–139, 2015. Disponível em: <http://dx.doi.org/10.1007/s10502-014-9236-y>. Acesso em: 16 set. 2017.

POOLE, A. H. The conceptual landscape of digital curation. *Journal of Documentation*, v. 72, n. 5, p. 961–986, 2016. Disponível em: <http://www.emeraldinsight.com/doi/10.1108/JD-10-2015-0123>. Acesso em: 16 set. 2017.

POUCHARD, L. Revisiting the data lifecycle with big data curation. *International Journal of Digital Curation*, v. 10, n. 2, p. 176–192, 2015. Disponível em: <http://www.whitehouse.gov/administration/eop/ostp/grand-challenges>. Acesso em: 24 jul. 2018.

POWELL, A.; HEERY, R. *Digital Repositories Roadmap: looking forward*. [S.l.]: UKOLN; Eduserv Foundation, 2006. Disponível em: <http://old.arl.org/resources/pubs/br/br226/br226ir.shtml>.

PRICE, D. J. DE S. *Little science, big science... and beyond*. New York: Columbia University Press, 1986.

PRÍNCIPE, P. Gestão de dados científicos: o papel das bibliotecas. out. 2013, São Paulo: [s.n.], out. 2013.

PRÍNCIPE, P.; SARAIVA, R. Serviços para suporte à gestão de dados científicos na UMINHO: plano de intervenção dos SDUM. CONGRESSO NACIONAL DE BIBLIOTECÁRIOS, ARQUIVISTAS E DOCUMENTALISTAS, 12., 2015, Évora. *Actas [...]*. [Lisboa]: Associação Portuguesa de Bibliotecários, Arquivistas e Documentalistas, 2015. p. 1–3. Disponível em:

http://www.bad.pt/publicacoes/index.php/congressosbad/article/view/1362/pdf_20. Acesso em: 11 set. 2017.

PRYOR, G. A patchwork of change. In: PRYOR, G.; JONES, S.; WHYTE, A. (org.). *Delivering research data management services: fundamentals of good practice*. London: Facet Publishing, 2014a. p. 1–19.

PRYOR, G. Preface. In: PRYOR, G.; JONES, S.; WHYTE, A. (org.). *Delivering research data management services: fundamentals of good practice*. London: Facet Publishing, 2014b. p. vii–viii.

PRYOR, G.; JONES, S.; WHYTE, A. (org.). *Delivering Research Data Management Services: fundamentals of good practice*. London: Facet Publishing, 2014.

PYTHON SOFTWARE FOUNDATION. *Welcome to Python.org*. Python.... [S.l.], 2017. Disponível em: <https://www.python.org/>. Acesso em: 5 out. 2017.

QIN, J. On the horizon of eScience Librarianship. *Journal of eScience Librarianship*, v. 1, n. 1, p. 2–3, 2012. Disponível em: <http://escholarship.umassmed.edu/jeslib/vol1/iss1/2/>. Acesso em: 5 out. 2017.

QIN, J. Infrastructure, standards, and policies for Research Data Management. In: INTERNATIONAL CONFERENCE ON COOPERATION AND PROMOTION OF INFORMATION RESOURCES IN SCIENCE AND TECHNOLOGY (COINFO), 8., 2013, Beijing. *The Proceedings of COINFO 2013*. Beijing: Science Press, 2013. p. 214–219. Disponível em: <https://surface.syr.edu/cgi/viewcontent.cgi?referer=https://www.google.com.br/&httpsredir=1&article=1164&context=istpub>. Acesso em: 8 maio 2018.

QIN, J.; CROWSTON, K.; KIRKLAND, A. Pursuing best performance in Research Data Management by using the capability maturity model and rubrics. *Journal of eScience Librarianship*, v. 6, n. 2, p. e1113 1-16, 2017. Disponível em: <http://escholarship.umassmed.edu/jeslib:%0Ahttps://doi.org/10.7191/jeslib.2017.1113>. Acesso em: 14 fev. 2019.

RABOIN, R.; REZNIK-ZELLEN, R.; SALO, D. Forging new service paths: institutional approaches to providing research data management services. *Journal of eScience Librarianship*, v.1, n.3, p. 134-147, 2012. Disponível em: <http://escholarship.umassmed.edu/jeslib/vol1/iss3/2/>. Acesso em: 29 maio 2018.

READ, K. B. et al. Starting the data conversation: informing data services at an academic health sciences library. *Journal of the Medical Library Association : JMLA*, v. 103, n. 3, p. 131–135, 2015. Disponível em: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4511052/>. Acesso em: 30 jul. 2018.

R FOUNDATION. *R: The R Project for Statistical Computing*. [S.l.]: R Foundation, [2017]. Disponível em: <https://www.r-project.org/>. Acesso em: 5 out. 2017.

REED, R. Diving into data: planning a research data management event. *Journal of eScience Librarianship*, v. 4, n. 1, p. e1071 1-6, 2015. Disponível em: <http://escholarship.umassmed.edu/jeslib/vol4/iss1/5/>. Acesso em: 16 jul. 2017.

REZNIK-ZELLEN, R.; ADAMICK, J.; MCGINTY, S. Tiers of research data support services. *Journal of eScience Librarianship*, v. 1, n. 1, p. 27–35, 2012. Disponível em: <http://escholarship.umassmed.edu/jeslib/vol1/iss1/5/>. Acesso em: 29 maio 2018.

RIBEIRO, C.; FERNANDES, M. E. M. Data Curation at U.Porto: identifying current practices across disciplinary domains. *IASSIST Quarterly*, v. 35, n. 4, p. 14–17, 2011. Disponível em:

[http://0-search.ebscohost.com.umaclib3.umac.mo/login.aspx?direct=true&db=lih&AN=76581083&site=eds-live%5Chttps://drive.google.com/file/d/0BwVuBK4FRW2zV3BMM3RRNmQyeTg/edit?usp=s](http://0-search.ebscohost.com/umaclib3.umac.mo/login.aspx?direct=true&db=lih&AN=76581083&site=eds-live%5Chttps://drive.google.com/file/d/0BwVuBK4FRW2zV3BMM3RRNmQyeTg/edit?usp=s)haring. Acesso em: 30 jul. 2018.

RICE, R.; SOUTHALL, J. *The Data Librarian's Handbook*. London: Facet Publishing, 2016.

RODRIGUES, E. *et al.* Os repositórios de dados científicos: estado da arte. Lisboa: RCAAP, 2010. Disponível em: <https://repositorio-aberto.up.pt/bitstream/10216/23806/2/44632.pdf>. Acesso em: 4 out. 2017.

RODRÍGUEZ BRAVO, B.; ALVITE DÍEZ, M. L. E-science and open access repositories in Spain. *OCLC Systems & Services: International digital library perspectives*, v. 23, n. 4, p. 363–371, 2007. Disponível em: <http://www.emeraldinsight.com/doi/10.1108/10650750710831501>. Acesso em: 5 jun. 2017.

ROSSI, T.; BEM, R. M.; LEHMKUHL, K. M. Suporte à pesquisa e gerenciamento de dados: proposta de concepção de serviço para Biblioteca Universitária da UFSC. In: SEMINÁRIO NACIONAL DE BIBLIOTECAS UNIVERSITÁRIAS, 19., 2016, Manaus. *Anais do SNBU*. Manaus: UFAM, 2016. p. 1–10. Disponível em: <http://www.periodicos.ufam.edu.br/anaisnibu/article/view/3282/2925>. Acesso em: 11 set. 2017.

SALES, L. F. *Integração semântica de publicações científicas e dados de pesquisa: proposta de modelo de publicação ampliada para a área de ciências nucleares*. 2014. 264f. Tese (Doutorado em Ciência da Informação) - Universidade Federal do Rio de Janeiro (UFRJ)/ Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT), Rio de Janeiro, 2014. Disponível em: http://repositorio.ibict.br/bitstream/123456789/874/1/LUANA_SALES_D.pdf. Acesso em: 20 set. 2017.

SALES, L. F.; SAYÃO, L. F. Ciberinfraestrutura de informação para a pesquisa: uma proposta de arquitetura para integração de repositórios e sistemas CRIS. *Informação e Sociedade*, v. 25, n. 3, p. 163–184, 2015. Disponível em: <http://periodicos.ufpb.br/ojs/index.php/ies/article/view/23998/14535>. Acesso em: 20 set. 2017.

SALES, L. F.; SAYÃO, L. F. A ciência invisível: relevando os dados da cauda longa da pesquisa. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 19., 2018, Londrina. [Anais...]. Londrina: Universidade Estadual de Londrina, 2018. 4180-4199. Disponível em: <http://enancib.marilia.unesp.br/index.php/XIXENANCIB/xixenancib/paper/viewFile/1538/1855>. Acesso em: 7 maio 2019.

SAYÃO, L. F. A ciência invisível: os dados da cauda longa da pesquisa científica. In: SEMINÁRIO DE SUPORTE À PESQUISA E GESTÃO DE DADOS CIENTÍFICOS: PANORAMA ATUAL E DESAFIOS, 1., 2017, Florianópolis. *Palestra*. Florianópolis: UFSC, 2017.

SAYÃO, L. F.; SALES, L. F. Curadoria digital: um novo patamar para preservação de dados digitais de pesquisa. *Informação e Sociedade*, v. 22, n. 3, p. 179–191, 2012. Disponível em: <https://periodicos.ufpb.br/index.php/ies/article/view/12224>. Acesso em: 20 set. 2017.

SAYÃO, L. F.; SALES, L. F. Dados de pesquisa: contribuição para o estabelecimento de um modelo de curadoria digital para o país. *Tendências da Pesquisa Brasileira em Ciência da Informação*, v. 6, n. 1, p. 1–26, 2013. Disponível em: <http://eprints.rclis.org/22562/2/102-459-1-PB.pdf>. Acesso em: 20 set. 2017.

SAYÃO, L. F.; SALES, L. F. *Guia de Gestão de Dados de Pesquisa para bibliotecários e pesquisadores*. Rio de Janeiro: CNEN/IEN, 2015. Disponível em: http://www.cnen.gov.br/images/CIN/PDFs/GUIA_DE_DADOS_DE_PESQUISA.pdf. Acesso em: 20 set. 2017.

SAYÃO, L. F.; SALES, L. F. Algumas considerações sobre os repositórios digitais de dados de pesquisa. *Informação & Informação*, v. 21, n. 2, p. 90–115, 2016. Disponível em: <http://www.uel.br/revistas/uel/index.php/informacao/article/view/27939>. Acesso em: 20 set. 2017.

SAYÃO, L. F.; SALES, L. F. Conectando a biblioteca de pesquisa à eScience: desafios e oportunidades. *CENDOC em Revista*, v. January, p. 27–37, 2017a. Disponível em: <http://www2.fab.mil.br/cendoc/images/conteudo/revista-cendoc.pdf>. Acesso em: 20 set. 2017.

SAYÃO, L. F.; SALES, L. F. Curadoria digital e dados de pesquisa. *AtoZ: novas práticas em informação e conhecimento*, v. 5, n. 2, p. 67, 2017b. Disponível em: <http://revistas.ufpr.br/atoz/article/view/49708>. Acesso em: 20 set. 2017.

SAYÃO, L. F.; SALES, L. F. *A ciência invisível: os dados da cauda longa da pesquisa científica*. [S.l.: s.n.], 2018. No prelo.

SCHMILLEN, H. *Library and Information Science Education and eScience : the current state of ALA Accredited MLS / MLIS Programs in preparing librarians and information professionals for eScience needs*. Denver: University of Denver, 2015. Disponível em: http://digitalcommons.du.edu/lis_capstone/1/. Acesso em: 6 jun. 2017.

SEMLER, A. R.; PINTO, A. L.; VIANNA, W. B. E-science: an epistemological analysis based on the philosophy of technology. *IFLA Journal*, v. 43, n. 2, p. 198–209, 2017. Disponível em: <http://journals.sagepub.com/doi/10.1177/0340035216678235>. Acesso em: 6 jun. 2017.

SESARTIC, A.; TÖWE, M. Research Data Services at ETH-Bibliothek. *IFLA Journal*, v. 42, n. 4, p. 284–291, 2016. Disponível em: <http://journals.sagepub.com/doi/10.1177/0340035216674971>. Acesso em: 30 jul. 2018.

SHEARER, K. *Comprehensive brief on research data management policies*. [S.l.: s.n.], 2015. Disponível em: <http://docplayer.net/17594465-Comprehensivebrief-on-research-data-management-policies.html>. Acesso em: 9 maio 2018.

SI, L. *et al.* Investigation and analysis of research data services in university libraries. *The Electronic Library*, v. 33, n. 3, p. 417–449, 2015. Disponível em: <http://www.emeraldinsight.com/doi/10.1108/EL-07-2013-0130>. Acesso em: 30 jul. 2018.

SOEHNER, C.; STEEVES, C.; WARD, J. *E-Science and data support services: a study of ARL member institutions*. Washington, D.C.: ARL, 2010. Disponível em: http://www.arl.org/bm~doc/escience_report2010.pdf. Acesso em: 07 fev. 2017.

STANFORD UNIVERSITY. *LOCKSS: Lots of Copies Keep Stuff Safe*. Stanford, CA: Stanford University, [2017]. Disponível em: <https://www.lockss.org/>. Acesso em: 5 out. 2017.

STARR, J. Information politics: the story of an emerging metadata standard. *First Monday*, v. 8, n. 7, 2003. Disponível em: <http://firstmonday.org/ojs/index.php/fm/article/view/1065/985>. Acesso em: 10 maio 2018.

STEWART, M. D.; LEWIS, B. R. A comprehensive analysis of marketing journal rankings. *Journal of Marketing Education*, v. 32, n. 1, p. 75–92, 2010. Disponível em: <http://journals.sagepub.com/doi/10.1177/0273475309344804>. Acesso em: 5 out. 2017.

SWANSON, J.; RINEHART, A. K. Data in context: using case studies to generate a common understanding of data in academic libraries. *Journal of Academic Librarianship*, v. 42, n. 1, p. 97–101, 2016. Disponível em: <http://dx.doi.org/10.1016/j.acalib.2015.11.005>. Acesso em: 2 set. 2017.

SZIGETI, K.; WHEELER, K. Science and technology resources on the internet: essential readings in e-science. *Issues in Science and Technology Librarianship*, v. 64, p. 1–10, 2011. Disponível em: <http://www.istl.org/11-winter/internet2.html>. Acesso em: 5 out. 2017.
TABLEAU Software. Seattle: Tableau Software, [2017]. Disponível em: <https://www.tableau.com/>. Acesso em: 5 out. 2017.

TENOPIR, C. et al. Academic librarians and research data services: preparation and attitudes. *IFLA Journal*, v. 39, n. 1, p. 70–78, 2013. Disponível em: <http://journals.sagepub.com/doi/10.1177/0340035212473089>. Acesso em: 2 set. 2017.

TENOPIR, C. et al. Research data management services in academic research libraries and perceptions of librarians. *Library & Information Science Research*, v. 36, n. 2, p. 84–90, 2014. Disponível em: <http://dx.doi.org/10.1016/j.lisr.2013.11.003>. Acesso em: 14 fev. 2019.

TENOPIR, C. et al. Research data services in academic libraries: data intensive roles for the future. *Journal of eScience Librarianship*, v. 4, n. 2, p. e1085 1-21, 2015. Disponível em: <http://escholarship.umassmed.edu/jeslib/vol4/iss2/4/>. Acesso em: 29 jun. 2017.

TENOPIR, C. et al. Research data services in european and north american libraries: Current offerings and plans for the future. *Proceedings of the Association for Information Science and Technology*, v. 53, n. 1, p. 1–6, 2016. Disponível em: <http://doi.wiley.com/10.1002/pra2.2016.14505301129>. Acesso em: 20 ago. 2018.

TENOPIR, C. et al. Research data services in European Academic Research Libraries. *LIBER QUARTERLY*, v. 27, n. 1, p. 23–44, 2017. Disponível em: <https://www.liberquarterly.eu/articles/10.18352/lq.10180/>. Acesso em: 10 ago. 2017.

TENOPIR, C.; BIRCH, B.; ALLARD, S. *Academic libraries and research data services: current practices and plans for the future*. [S.l.]: ACRL, 2012. Disponível em: http://www.ala.org/acrl/sites/ala.org.acrl/files/content/publications/whitepapers/Tenopir_Birch_Allard.pdf. Acesso em: 29 jun. 2017. (Whipe Paper).

THORPE, A. Environmental eScience. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, v. 367, n. 1890, p. 801–802, 2009. Disponível em: <http://rsta.royalsocietypublishing.org/cgi/doi/10.1098/rsta.2008.0254>. Acesso em: 7 jun. 2017.

TJALSMA, H.; ROMBOUTS, J. *Selection of of research data: guidelines for appraising and selecting research data*. Den Haag: Data Archiving and Networked Services, 2011. v. 6. Disponível em: <http://www.dans.knaw.nl/content/categorieen/publicaties/dans-studies-digital-archiving-6>. (DANS Studies in Digital Archiving).

TRIPATHI, M.; SHUKLA, A.; SONKAR, S. K. Research data management practices in university libraries: a study. *DESIDOC Journal of Library & Information Technology*, v. 37, n. 6, p. 417, 2017. Disponível em: <http://publications.drdo.gov.in/ojs/index.php/djlit/article/view/11336>. Acesso em: 30 jul. 2018.

UK DATA SERVICE. *Research data lifecycle*. [S.l.], [2018]. Disponível em: <https://www.ukdataservice.ac.uk/manage-data/lifecycle>. Acesso em: 27 jul. 2018.

UNITED STATES ENVIRONMENTAL PROTECTION AGENCY (USEPA). *Learn About Data Standards*. Washington, D.C.: USEPA, 2018. Disponível em: <https://www.epa.gov/data->

standards/learn-about-data-standards. Acesso em: 14 maio 2018.

UNITED STATES GEOLOGICAL SURVEY (USGS). *USGS Data Management: Data Standards*. Reston: USGS, 2018a. Disponível em: <https://www2.usgs.gov/datamanagement/plan/datastandards.php>. Acesso em: 14 maio 2018.

UNITED STATES GEOLOGICAL SURVEY (USGS). *USGS Data Management: Data Lifecycle Overview*. Reston: USGS, 2018b. Disponível em: <https://www2.usgs.gov/datamanagement/why-dm/lifecycleoverview.php>. Acesso em: 19 jul. 2018.

UNIVERSITY OF CALIFORNIA (UC). California Digital Library. ARK. Oakland: UC, 2019. Disponível em: <http://n2t.net/e/ark_ids.html>. Acesso em: 13 fev. 2019.

VAN DE SOMPEL, H. *et al.* Rethinking scholarly communication. *D-Lib Magazine*, v. 10, n. 9, 2004. Disponível em: <http://www.dlib.org/dlib/september04/vandesompel/09vandesompel.html>. Acesso em: 20 mar. 2018.

VERETENNIKOVA, N. *et al.* E-Science: new paradigms, system integration and scientific research organization. Lviv, Ukraine: IEEE, 2015. p. 76–81. Disponível em: <http://escholarship.umassmed.edu/jeslib/vol1/iss1/6/>. Acesso em: 18 ago. 2017.

VITALE, C. R. H. Is research reproducibility the new data management for libraries. *Bulletin of the Association for Information Science & Technology*, v. 42, n. 3, p. 38–41, 2016. Disponível em: <https://onlinelibrary.wiley.com/doi/full/10.1002/bul2.2016.1720420313>. Acesso em: 30 jul. 2018.

VLAEMINCK, S. Data management in scholarly journals and possible roles for libraries: some insights from EDaWaX. *LIBER Quarterly*, v. 23, n. 1, p. 48–79, 2013. Disponível em: <https://www.liberquarterly.eu/article/10.18352/lq.8082/>. Acesso em: 18 ago. 2017.

WALLIS, J. C.; ROLANDO, E.; BORGMAN, C. L. If we share data, will anyone use them. Data sharing and reuse in the long tail of science and technology. *PLoS ONE*, v. 8, n. 7, p. e67332 1-17, 2013. Disponível em: <http://dx.plos.org/10.1371/journal.pone.0067332>. Acesso em: 22 set. 2017.

WANG, M. Supporting the research process through expanded library data services. *Program: electronic library and information systems*, v. 47, n. 3, p. 282–303, 2013. Disponível em: <http://www.emeraldinsight.com/doi/10.1108/PROG-04-2012-0010>. Acesso em: 22 set. 2017.

WATSON, P.; TREFETHEN, A.; VANDER MEER, E. e-Science: past, present and future I. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, v. 368, n. 1925, p. 3759–3760, 2010. Disponível em: <http://rsta.royalsocietypublishing.org/cgi/doi/10.1098/rsta.2010.0177>. Acesso em: 22 set. 2017.

WELCOME, Scientific Data. *Nature*, v. 509, n. 7502, p. 534–534, 2014. Disponível em: <http://www.nature.com/news/welcome-scientific-data-1.15293>. Acesso em: 8 set. 2017.

WELLER, T.; MONROE-GULICK, A. Differences in the data practices, challenges, and future needs of graduate students and faculty members. *Journal of eScience Librarianship*, v. 4, n. 1, p. 21070 1-12, 2015. Disponível em: <http://escholarship.umassmed.edu/jeslib/vol4/iss1/2/>. Acesso em: 18 out. 2017.

WELLER, T.; MONROE-GULICK, A. Understanding methodological and disciplinary differences in the data practices of academic researchers. *Library Hi Tech*, v. 32, n. 3, p.

467–482, 2014. Disponível em: <http://www.emeraldinsight.com/doi/10.1108/LHT-02-2014-0021>. Acesso em: 18 out. 2017.

WHITMIRE, A. L.; BOOCK, M.; SUTTON, S. C. Variability in academic research data management practices. *Program: electronic library and information systems*, v. 49, n. 4, p. 382–407, 2015. Disponível em: <http://www.emeraldinsight.com/doi/10.1108/PROG-02-2015-0017>. Acesso em: 18 out. 2017.

WHYTE, A.; ALLARD, S. *How to discover requirements for research data management services*. Edinburgh: DCC, 2014. Disponível em: http://www.dcc.ac.uk/sites/default/files/documents/publications/DCC_Howto_Discover_Requirements.pdf. Acesso em: 19 jun. 2018. (DCC How-to Guides).

WHYTE, A.; JONATHAN, T. *Making the case for research data management*. DCC Briefing Papers. Edinburgh: DCC, 2011. Disponível em: [http://www.dcc.ac.uk/sites/default/files/documents/publications/Making the case.pdf](http://www.dcc.ac.uk/sites/default/files/documents/publications/Making%20the%20case.pdf). Acesso em: 8 maio 2018.

WHYTE, A.; WILSON, A. *How to Appraise & Select Research Data for Curation*. [S.l.]: DCC, 2010. Disponível em: <http://www.dcc.ac.uk/resources/how-guides/appraise-select-data>. Acesso em: 8 set. 2017.

WILEY, C.; MISCHO, W. H. Data management practices and perspectives of Atmospheric Scientists and Engineering Faculty. *Issues in Science and Technology Librarianship*, n. Fall, 2016. Disponível em: <http://istl.org/16-fall/refereed3.html#20>. Acesso em: 17 out. 2017.

WILLIS, C.; GREENBERG, J.; WHITE, H. Analysis and synthesis of metadata goals for scientific data. *Journal of the American Society for Information Science and Technology*, v. 63, n. 8, p. 1505–1520, 2012. Disponível em: <http://doi.wiley.com/10.1002/asi.22683>. Acesso em: 8 set. 2017.

WILSON, J. *et al.* Developing infrastructure for research data management at the University of Oxford. *Ariadne*, n. 65, 2010. Disponível em: <http://www.ariadne.ac.uk/issue65/wilson-et-al>. Acesso em: 18 abr. 2018.

WILSON, J. A. J. *et al.* An institutional approach to developing research data management infrastructure. *International Journal of Digital Curation*, v. 6, n. 2, p. 274–287, 2011. Disponível em: <http://ijdc.net/index.php/ijdc/article/view/198>. Acesso em: 18 abr. 2018.

WRIGHT, M. *et al.* Connecting digital libraries to eScience: the future of scientific scholarship. *International Journal on Digital Libraries*, v. 7, n. 1–2, p. 1–4, 2007. Disponível em: <https://link.springer.com/article/10.1007/s00799-007-0030-9>. Acesso em: 8 set. 2017.

YAKEL, E. Digital curation. *OCLC Systems & Services: International digital library perspectives*, v. 23, n. 4, p. 335–340, 2007. Disponível em: <http://www.emeraldinsight.com/doi/10.1108/10650750710831466>. Acesso em: 16 set. 2017.

YOON, A.; SCHULTZ, T. Research data management services in academic libraries in the US: a content analysis of libraries' websites. *College & Research Libraries*, v. 78, n.7, p. 1–19, 2017. Disponível em: <http://crl.acrl.org/index.php/crl/article/view/16788>. Acesso em: 28 maio 2018.

YU, F.; DEUBLE, R.; MORGAN, H. Designing research data management services based on the research lifecycle: a consultative leadership approach. *Journal of the Australian Library and Information Association*, v. 663, n. 3, p. 287–298, 2017. Disponível em: <http://www.tandfonline.com/action/journalInformation?journalCode=ualj21%0Ahttp://dx.doi.org/10.1080/24750158.2017.1364835%0Ahttps://doi.org/10.1080/24750158.2017.1364835>.

Acesso em: 4 fev. 2019.

YU, H. H. The role of academic libraries in research data service (RDS) provision: opportunities and challenges. *The Electronic Library*, v. 35, n. 4, p. 783–797, 3 jul. 2017. Disponível em: <http://www.emeraldinsight.com/doi/10.1108/09574090910954864>. Acesso em: 18 ago. 2017.

ZHU, H.; LEE, Y. W.; ROSENTHAL, A. S. Data standards challenges for interoperable and quality data. *Journal of Data and Information Quality*, v. 7, n. 1–2, p. 1–3, 2016. Disponível em: <http://dl.acm.org/citation.cfm?doid=2888577.2903723>. Acesso em: 14 maio 2018.

ZHU, H.; WU, H. Quality of data standards: framework and illustration using XBRL taxonomy and instances. *Electronic Markets*, v. 21, n. 2, p. 129–139, 2011. Disponível em: <http://link.springer.com/10.1007/s12525-011-0060-4>. Acesso em: 14 maio 2018.

ZORICH, D. M. Data management: managing electronic information: Data curation in museums. *Museum Management and Curatorship*, v. 14, n. 4, p. 430–432, 1995. Disponível em: <http://www.sciencedirect.com/science/article/pii/0260477996846905>. Acesso em: 16 set. 2017.

APÊNDICES

APÊNDICE A – PADRÕES DE METADADOS

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

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

Fonte: Elaborado pela autora com base em DCC (2018b).

APÊNDICE B – EXTENSÕES

Área	Extensão	Informações
Geral	AGLS Metadata Profile	http://www.agls.gov.au/
Geral	Asset Description Metadata Schema (ADMS)	http://www.w3.org/TR/vocab-adms/
Geral	Dryad Metadata Application Profile	http://wiki.datadryad.org/Metadata_Profile
Geral	GSIM - Generic Statistical Information Model	http://www1.unece.org/stat/platform/display/metis/Generic+Statistical+Information+Model
Geral	OpenAIRE Guidelines for publication repositories, data archives and CRIS systems	https://guidelines.openaire.eu/
Geral	Tabular Data Package	http://www.dcc.ac.uk/resources/subject-areas/general-research-data
Biologia	ABCDDNA	http://www.tdwg.org/standards/640/
Biologia	Apple Core	http://code.google.com/p/applecore/wiki/Introduction
Biologia	Darwin Core Geospatial Extension	http://wiki.tdwg.org/twiki/bin/view/DarwinCore/GeospatialExtension
Biologia	DwC Germplasm	http://code.google.com/p/darwincore-germplasm/
Biologia	EDMED Metadata Profile	https://www.bodc.ac.uk/data/information_and_inventories/edmed/
Biologia	FGDC/CSDGM Biological Data Profile	http://www.fgdc.gov/standards/projects/FGDC-standards-projects/metadata/biometadata/biodatap.pdf
Biologia	GBIF Metadata Profile	http://rs.gbif.org/schema/eml-gbif-profile/
Biologia	HISPID - Herbarium Information Standards and Protocols for Interchange of Data	http://www.rbg.vic.gov.au/wiki/hiscom/index.php/HISPID_5
Biologia	ISA-TAB Nano	https://wiki.nci.nih.gov/display/ICR/ISA-TAB-Nano#ISA-TAB-Nano-title
Biologia	isaconfig-diXa	https://bitbucket.org/kanerae/isaconfig-dixa
Biologia	MIBBI Portal	http://mibbi.sourceforge.net/portal.shtml
Biologia	OME-TIFF - Open Microscopy Environment TIFF	http://www.openmicroscopy.org/site/support/file-formats/ome-tiff
Biologia	SNRNASM ISA-Tab	http://snrnasm.bio.unc.edu/
Biologia	VarioML	http://www.ncbi.nlm.nih.gov/pubmed/23031277
Ciências da Terra	ABCDEFGF - Access to Biological Collection Databases Extended for Geosciences	http://www.geocase.eu/efg
Ciências da Terra	AGRIS Application Profile	http://www.fao.org/docrep/008/ae909e/ae909e00.htm
Ciências da Terra	ANZLIC Metadata Profile	http://www.anzlic.gov.au/resources/metadata
Ciências da Terra	Cruise Summary Reports	http://www.seadatanet.org/Metadata/CSR
Ciências da Terra	CSDGM Extensions for Remote Sensing Metadata	http://www.fgdc.gov/standards/projects/FGDC-standards-projects/csdgm_rs_ex/MetadataRemoteSensingExtens.pdf
Ciências da Terra	Darwin Core Geospatial Extension	http://wiki.tdwg.org/twiki/bin/view/DarwinCore/GeospatialExtension
Ciências da Terra	EDMED Metadata Profile	https://www.bodc.ac.uk/data/information_and_inventories/edmed/
Ciências da Terra	INSPIRE Metadata Regulation	http://inspire.ec.europa.eu/
Ciências da Terra	ISO 19115 North American Profile	http://www.fgdc.gov/standards/projects/incits-I1-standards-projects/NAP-Metadata
Ciências da Terra	ISO 19115-2 - Imagery and gridded data	http://www.iso.org/iso/catalogue_detail.htm?csnumber=39229
Ciências da Terra	Marine Community Profile	http://www.aodc.org.au/index.php?id=37

Ciências da Terra	Shoreline Metadata Profile	http://www.fgdc.gov/standards/projects/FGDC-standards-projects/metadata/shoreline-metadata/
Ciências da Terra	UK AGMAP - Academic Geospatial Metadata Application Profile	http://www.gogeo.ac.uk/gogeo/metadata/agmap.htm
Ciências da Terra	UK Gemini	http://www.agi.org.uk/uk-gemini/
Ciências da Terra	USGIN ISO profile	http://repository.stategeothermaldata.org/repository/resource/98ddf901b9782a25982e01af3b0bda50/
Ciências da Terra	WaterML	http://www.opengeospatial.org/standards/waterml
Ciências da Terra	WMO Core Metadata Profile	http://www.wmo.int/pages/prog/www/WIS/metadata_en.html
Ciências Físicas	eBank UK Metadata Application Profile	http://ecrystals.chem.soton.ac.uk/
Ciências Físicas	FITS World Coordinate System (WCS)	http://fits.gsfc.nasa.gov/fits_wcs.html
Ciências Físicas	IMPEX Data Model	http://meetingorganizer.copernicus.org/EPSC2012/EPSC2012-360.pdf
Ciências Físicas	Resource Metadata for the Virtual Observatory	http://www.ivoa.net/documents/latest/RM.html
Ciências Físicas	TIDCC - Towards an International Data Commons for Crystallography	http://wiki.ecrystals.chem.soton.ac.uk/images/9/9d/ECrystals-WP4-PM-Final.pdf
Ciências Físicas	MO Core Metadata Profile	http://www.wmo.int/pages/prog/www/WIS/metadata_en.html
Ciências Sociais e Humanidades	CARARE metadata schema	http://carare.eu/dut/Resources/CARARE-Documentation/CARARE-metadata-schema
Ciências Sociais e Humanidades	CESSDA MLI - Council of European Social Science Data Archives Minimum Level of Information	http://www.cessda.org/sharing/managing/3/
Ciências Sociais e Humanidades	GESMES/TS - GEneric Statistical MESsage for Time Series	http://www.ecb.int/stats/services/sdmx/html/index.en.html

Fonte: Elaborado pela autora com base em DCC (2018b).

APÊNDICE C – LISTA DE ARTIGOS

Autor	Nome	Periódico	Ano
Mary Anne Kennan; Sheila Corral; Waseem Afzal	"Making space" in practice and education: research support services in academic libraries	Library Management	2014
Kevin Read; Fred Willie Zametkin LaPolla	A new hat for librarians: providing REDCap support to establish the library as a central data hub	Journal of the Medical Library Association	2018
Carol Tenopir <i>et al</i>	Academic librarians and research data services: preparation and attitudes	IFLA Journal	2012
Elisha Chiware; Zanele Mathe	Academic libraries' role in Research Data Management Services: a South African perspective	South African Journal of Library and Information Science	2015
Luis Fernando Sayão; Luana Farias Sales	Algumas considerações sobre os repositórios digitais de dados de pesquisa	Informação & Informação	2016
Andrew Creamer; Andrew Creamer; Javier Crespo	An Assessment of Needed Competencies to Promote the Data Curation and Management Librarianship of Health Sciences and Science and Technology Librarians in New England	Journal of eScience Librarianship	2012
Elizabeth A. Berman	An Exploratory Sequential Mixed Methods Approach to Understanding Researchers' Data Management Practices at UVM: Findings from the Qualitative Phase	Journal of eScience Librarianship	2017
Elizabeth A. Berman	An Exploratory Sequential Mixed Methods Approach to Understanding Researchers' Data Management Practices at UVM: Findings from the Quantitative Phase	Journal of eScience Librarianship	2017
Elizabeth A. Berman	An Exploratory Sequential Mixed Methods Approach to Understanding Researchers' Data Management Practices at UVM: Integrated Findings to Develop Research Data Services	Journal of eScience Librarianship	2017
Soohyung Joo; Sujin Kim; Youngseek Kim	An exploratory study of health scientists' data reuse behaviors: examining attitudinal, social, and resource factors	Aslib Journal of Information Management	2017
James A. J. Wilson <i>et al</i>	An Institutional Approach to Developing Research Data Management Infrastructure	International Journal of Digital Curation	2011
Tibor Koltay	Are you ready? Tasks and roles for academic libraries in supporting Research 2.0	New Library World	2016
Christie A. Wiley	Assessing Research Data Deposits and Usage Statistics within IDEALS	Journal of eScience Librarianship	2017
Megan M. Bresnahan; Andrew M. Johnson	Assessing scholarly communication and research data training needs	Reference Services Review	2013
Julie Goldman; Donna Kafel; Elaine R. Martin	Assessment of Data Management Services at New England Region Resource Libraries	Journal of eScience Librarianship	2015
Leslie M. Delsertone	At the Watershed: Preparing for Research Data Management and Stewardship at the University of Minnesota Libraries	Library Trends	2008
Sebastian Karcher; Dessislava Kirilova; Nicholas Weber	Beyond the matrix: Repository services for qualitative data	IFLA Journal	2016
Sheila Corral; Mary Anne Kennan; Waseem Afzal	Bibliometrics and Research Data Management Services: Emerging Trends in Library Support for Research	Library Trends	2013
Laura Gordon-Murnane	Big data: a big opportunity for librarians	Online	2012
Michele P. Claibourn	Bigger on the inside: building Research Data Services at the University of Virginia	Insights	2015
Jamie Wittenberg; Mary Elings	Building a Research Data Management Service at the University of California, Berkeley: A tale of collaboration	IFLA Journal	2017
Gareth Knight	Building a research data management service for the London school of hygiene & tropical medicine	Program: electronic library and information systems	2015

Cheryl A. Thompson	Building Data Expertise into Research Institutions: Preliminary Results	Proceedings of the Association for Information Science and Technology	2015
Heather L. Coates	Building Data Services From the Ground Up: Strategies and Resources	Journal of eScience Librarianship	2014
Suzanna Conrad <i>et al</i>	Building professional development opportunities in data services for academic librarians	IFLA Journal	2017
Katherine G. Akers <i>et al</i>	Building Support for Research Data Management: Biographies of Eight Research Universities	International Journal of Digital Curation	2014
Donna Kafel; Andrew T. Creamer; Elaine R. Martin	Building the New England Collaborative Data Management Curriculum	Journal of eScience Librarianship	2014
Stephanie Wright <i>et al</i>	Collaboration and tension between institutions and units providing data management support	Bulletin of the Association for Information Science and Technology	2014
Karen Hogenboom; Tom Teper; Lynn Wiley	Collecting Small Data	Research Library Issues	2011
Luís Fernando Sayão; Luana Farias Sales	Conectando a biblioteca de pesquisa à eScience: desafios e oportunidades	CENDOC em Revista	2017
Thomas Parsons	Creating a Research Data Management Service	International Journal of Digital Curation	2013
Hans Jørn Nielsen; Birger Hjørland	Curating research data: the potential roles of libraries and information professionals	Journal of Documentation	2014
Shiyan Ou; Yu Zhou	Current Status of Scientific Data Curation Research and Practices in Mainland China	Libres	2016
Anna Gold	Cyberinfrastructure, Data, and Libraries, Part 1: A Cyberinfrastructure Primer for Librarians	D-Lib Magazine	2007
Anna Gold	Cyberinfrastructure, Data, and Libraries, Part 2: Libraries and the Data Challenge: Roles and Actions for Libraries	D-Lib Magazine	2007
Luís Fernando Sayão; Luana Farias Sales	Dados abertos de pesquisa: ampliando o conceito de acesso livre	Revista Eletrônica de Comunicação, Informação e Inovação em Saúde	2014
Cristina Ribeiro; Maria Eugénia Matos Fernandes	Data Curation at U.Porto: Identifying current practices across disciplinary domains	IASSIST QuArTerly	2011
Merinda McLure <i>et al</i>	Data Curation: A Study of Researcher Practices and Needs	portal: Libraries and the Academy	2014
Alisa Surkis <i>et al</i>	Data Day to Day: building a community of expertise to address data skills gaps in an academic medical center	Journal of the Medical Library Association	2017
Juleah Swanson; Amanda K. Rinehart	Data in context: Using case studies to generate a common understanding of data in academic libraries	The Journal of Academic Librarianship	2016
Katharine V. Macy	Data information literacy instruction in Business and Public Health: comparative case studies	IFLA Journal	2016
Stuart Macdonald; Luis Martinez-Urbe	Data librarianship: a gap in the market	update	2008
Tibor Koltay	Data literacy for researchers and data librarians	Journal of Librarianship and Information Science	2017
Lisa M. Federer; Ya-Ling Lu; Douglas J. Joubert	Data literacy training needs of biomedical researchers	Journal of the Medical Library Association	2016
Tibor Koltay	Data literacy: in search of a name and identity	Journal of Documentation	2015
VIRGIL E. VARVEL, JR; YI SHEN	Data Management Consulting at The Johns Hopkins University	New Review of Academic Librarianship	2013
Sven Vlaeminck	Data Management in Scholarly Journals and Possible Roles for Libraries: some Insights from EDaWaX	Liber Quarterly	2013

Andrew M. Johnson; Shelley Knuth	Data Management Plan Requirements for Campus Grant Competitions: Opportunities for Research Data Services Assessment and Outreach	Journal of eScience Librarianship	2016
Jessica Adamick; Rebecca C. Reznik- Zellen; Matt Sheridan	Data Management Training for Graduate Students at a Large Research University	Journal of eScience Librarianship	2012
Brian Westra	Data Services for the Sciences: a needs assessment	Ariadne	2010
Eleanor J. Read	Data services in academic libraries: assessing needs and promoting services	Reference and User Services Quarterly	2007
Don MacMillan	Data Sharing and Discovery: What Librarians Need to Know	The Journal of Academic Librarianship	2014
Kim Jeonghyun	Data sharing and its implications for academic libraries	New Library World	2013
Shawn W. Nicholson; Terrence B. Bennett	Data Sharing: Academic Libraries and the Scholarly Enterprise	portal: Libraries and the Academy	2011
Donna Ellen Frederick	Data, Open Science and libraries – The Data Deluge Column	Library Hi Tech News	2016
Laura Rueda; Martin Fenner; Patricia Cruse	DataCite: Lessons Learned on Persistent Identifiers for Research Data	International Journal of Digital Curation	2016
Karen Antell <i>et al</i>	Dealing with Data: Science Librarians' Participation in Data Management at Association of Research Libraries Institutions	College & Research Libraries	2014
Lisa Federer	Defining data librarianship: a survey of competencies, skills, and training	Journal of the Medical Library Association	2018
Dianne Dietrich <i>et al</i>	De-Mystifying the Data Management Requirements of Research Funders	Issues in Science and Technology Librarianship	2012
Fei Yu; Rebecca Deuble; Helen Morgan	Designing Research Data Management Services Based on the Research Lifecycle – A Consultative Leadership Approach	Journal of the Australian Library and Information Association	2017
Jacob Carlson <i>et al</i>	Determining Data Information Literacy Needs: A Study of Students and Research Faculty	portal: Libraries and the Academy	2011
Jeff Moon	Developing a Research Data Management Service : a Case Study	Partnership: The Canadian Journal of Library and Information Practice and Research	2014
Layne M. Johnson; John T. Butler; Lisa R. Johnston	Developing E-science and Research Services and Support at the University of Minnesota Health Sciences Libraries	Journal of Library Administration	2012
James A. J. Wilson <i>et al</i>	Developing Infrastructure for Research Data Management at the University of Oxford	Ariadne	2010
Rebecca A. Brown; Malcolm Wolski; Joanna Richardson	Developing new skills for research support librarians	The Australian Library Journal	2015
Laure Perrier; Leslie Barnes	Developing research data management services and support for researchers: A mixed methods study	Partnership: The Canadian Journal of Library and Information Practice and Research	2018
K.T.L. Vaughan <i>et al</i>	Development of the research lifecycle model for library services	Journal of the Medical Library Association	2013
Andrew M. Cox <i>et al</i>	Developments in Research Data Management in Academic Libraries: Towards an Understanding of Research Data Service Maturity	Journal of the Association for Information Science and Technology	2017
Travis Weller; Amalia Monroe-Gulick	Differences in the Data Practices, Challenges, and Future Needs of Graduate Students and Faculty Members	Journal of eScience Librarianship	2015
Robert Fox	Digital libraries: the systems analysis perspective	OCLC Systems & Services: International digital library perspectives	2012
Claire Hamasu; Barb Jones; Betsy Kelly	Discussing “eScience and the Evolution of Library Services”	Journal of eScience Librarianship	2012
Robyn B. Reed	Diving into Data: Planning a Research Data Management Event	Journal of eScience Librarianship	2015

JoAnne Sparks <i>et al</i>	Embedding innovation for scholarly information and research	Library Management	2013
Sally A. Gore	e-Science and Data Management Resources on the Web	Medical Reference Services Quarterly	2011
Achim Osswald	E-science and information services: a missing link in the context of digital libraries	Online Information Review	2008
Tony Hey; Jessie Hey	e-Science and its implications for the library community	Library Hi Tech	2006
Hayes, Sheila	eScience. What Is This? And, What Does It Have to Do with Hospital Librarians?	Journal of Hospital Librarianship	2014
Gareth Cole	Establishing a Research Data Management Service at Loughborough University	International Journal of Digital Curation	2016
John Kaye; Rachel Bruce; Dom Fripp	Establishing a shared research data service for UK universities	Insights	2017
Regina Raboin; Rebecca C. Reznik-Zellen; Dorothea Salo	Forging New Service Paths: Institutional Approaches to Providing Research Data Management Services	Journal of eScience Librarianship	2012
Margaret H. Burnette; Sarah C. Williams; Heidi J. Imker	From Plan to Action: Successful Data Management Plan Implementation in a Multidisciplinary Project	Journal of eScience Librarianship	2016
Louise Patterson; Theo J.D. Bothma; Martie J. van Deventer	From planning to practice: an action plan for the implementation of research data management services in resource-constrained institutions	South African Journal of Library and Information Science	2018
Kristin A. Briney	Gaining Competency: Learning to Teach Data Visualization	Journal of eScience Librarianship	2018
Donna L. O'Malley	Gaining Traction in Research Data Management Support: A Case Study	Journal of eScience Librarianship	2014
Sarah C. Williams	Gathering Feedback from Early-Career Faculty: Speaking with and Surveying Agricultural Faculty Members about Research Data	Journal of eScience Librarianship	2013
Katherine G. Akers	Going beyond data management planning: Comprehensive research data services	College & Research Libraries News	2014
Thea P. Atwood; Patricia B. Condon; Julie Goldman	Grassroots Professional Development	Journal of eScience Librarianship	2017
Luana Farias Sales; Luís Fernando Sayão	Há Futuro para as Bibliotecas de Pesquisa no Ambiente de eScience	Informação & Tecnologia	2015
Tania P. Bardyn <i>et al</i>	Health Sciences Libraries Advancing Collaborative Clinical Research Data Management in Universities	Journal of eScience Librarianship	2018
Peter Linde <i>et al</i>	How can libraries and other academic stakeholders engage in making data open?	Information Services & Use	2014
Elizabeth Stephenson; Patti Schifter Caravello	Incorporating data literacy into undergraduate information literacy programs in the social sciences	Reference Services Review	2007
Christie Peters; Porcia Vaughn	Initiating Data Management Instruction to Graduate Students at the University of Houston Using the New England Collaborative Data Management Curriculum	Journal of eScience Librarianship	2014
Debra Hiom <i>et al</i>	Introducing Safe Access for Sensitive Data at the University of Bristol	International Journal of Digital Curation	2017
Li Si; Limei Zhou	Investigation and analysis of research data services in university libraries	The Electronic Library	2015
Cynthia R.H. Vitale	Is Research Reproducibility the New Data Management for Libraries?	Bulletin of the Association for Information Science and Technology	2016
Christine L. Borgman <i>et al</i>	Knowledge infrastructures in science: data, diversity, and digital libraries	International Journal on Digital Libraries	2015
Jennifer Muilenburg; Mahria Lebow; Joanne Rich	Lessons Learned From a Research Data Management Pilot Course at an Academic Library	Journal of eScience Librarianship	2014
Terence K. Huwe	Librarians and Data: Curator, Creator, or Both?	Online Searcher	2017
Samantha Searle <i>et al</i>	Librarians as partners in research data service development at Griffith University	Program: electronic library and information systems	2015

Shamin Renwick; Marsha Winter; Michelle Gill	Managing research data at an academic library in a developing country	IFLA Journal	2017
Gail Steinhart; Jian Qin	Mentoring for Emerging Careers in eScience Librarianship: An iSchool – Academic Library Partnership	Journal of eScience Librarianship	2012
Susan Hickson <i>et al</i>	Modifying researchers' data management practices: A behavioural framework for library practitioners	IFLA Journal	2016
Maira Murrieta Costa; Murilo Bastos da Cunha	O bibliotecário no tratamento de dados oriundos da e-science: considerações iniciais	Perspectivas em Ciência da Informação	2014
Luana Farias Sales; Luís Fernando Sayão	O impacto da curadoria digital dos dados de pesquisa na Comunicação Científica	Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação	2012
Fabiano Couto Corrêa	O papel dos bibliotecários na gestão de dados científicos	RDBCI: Revista Digital de Biblioteconomia e Ciência da Informação	2016
E. Verbaan; A.M. Cox	Occupational Sub-Cultures, Jurisdictional Struggle and Third Space: Theorising Professional Service Responses to Research Data Management	The Journal of Academic Librarianship	2014
Jake R. Carlson	Opportunities and Barriers for Librarians in Exploring Data: Observations from the Data Curation Profile Workshops	Journal of eScience Librarianship	2013
Christopher Eaker	Planning Data Management Education Initiatives: Process, Feedback, and Future Directions	Journal of eScience Librarianship	2014
Dong Joon Lee; Besiki Stvilia	PPactices of research data curation in institutional repositories: A qualitative view from repository staff	PLoS ONE	2017
Jian Qin; Kevin Crowston; Arden Kirkland	Pursuing Best Performance in Research Data Management by Using the Capability Maturity Model and Rubrics	Journal of eScience Librarianship	2017
Tibor Koltay	Research 2.0 and research data services in academic and research libraries: priority issues	Library Management	2017
Kevin Ashley	Research data and libraries: who does what	Insights	2012
Alisa Surkis; Kevin Read	Research data management	Journal of the Medical Library Association	2015
Andrew M. Cox; Stephen Pinfield	Research data management and libraries: Current activities and future priorities	Journal of Librarianship and Information Science	2014
Stephen Pinfield; Andrew M. Cox; Jen Smith	Research Data Management and Libraries: Relationships, Activities, Drivers and Influences	PLoS ONE	2014
Rosie Higman; Stephen Pinfield	Research data management and openness: the role of data sharing in developing institutional policies and practices	Program: electronic library and information systems	2015
Sarah Barbrow; Denise Brush; Julie Goldman	Research data management and services: resources for novice data librarians	College & Research Libraries News	2017
Michael Steeleworthy	Research Data Management and the Canadian Academic Library: An Organizational Consideration of Data Management and Data Stewardship	Partnership: the Canadian Journal of Library and Information Practice and Research	2014
Chris Awre <i>et al</i>	Research Data Management as a "wicked problem"	Library Review	2015
Alex Guindon	Research Data Management at Concordia University: A Survey of Current Practices	Feliciter	2014
Debra Hiom <i>et al</i>	Research data management at the University of Bristol Charting a course from project to service	Program: electronic library and information systems	2015
Joanna Ball	Research data management for libraries: getting started	Insights	2013
Pierre-Yves Burgi; Eliane Blumer; Basma Makhoul-Shabou	Research data management in Switzerland: National efforts to guarantee the sustainability of research outputs	IFLA Journal	2017

Joachim Schöpfel <i>et al</i>	Research data management in the French National Research Center (CNRS)	Data Technologies and Applications	2018
Arsev Umur Aydinoglu; Guleda Dogan; Zehra Taskin	Research data management in Turkey: perceptions and practices	Library Hi Tech	2017
Xia Liu; Ning Ding	Research data management in universities of central China: practices at Wuhan University Library	The Electronic Library	2016
Willow F. Dressel	Research Data Management Instruction for Digital Humanities	Journal of eScience Librarianship	2017
Manorama Tripathi; Archana Shukla; Sharad Kumar Sonker	Research data Management Practices in university libraries: a study	DESIDOC Journal of Library & Information Technology	2017
Constanze Curdt; Dirk Hoffmeister	Research data management services for a multidisciplinary, collaborative research project: Design and implementation of the TR32DB project database	Program: electronic library and information systems	2015
Ayoung Yoon; Teresa Schultz	Research Data Management Services in Academic Libraries in the US: A Content Analysis of Libraries' Websites	College & Research Libraries	2017
Carol Tenopir <i>et al</i>	Research data management services in academic research libraries and perceptions of librarians	Library & Information Science Research	2014
Elisha R.T. Chiware; Deborah Anne Becker	Research Data Management Services in Southern Africa: A Readiness Survey of Academic and Research Libraries	African Journal of Library, Archives and Information Science	2018
Gary Brewerton	Research data management: a case study	Ariadne	2015
Bhojaraju Gunjal; Panorea Gaitanou	Research Data Management: A proposed framework to boost research in Higher Educational Institutes	IASSIST Quarterly	2017
Bethany Latham	Research Data Management: Defining Roles, Prioritizing Services, and Enumerating Challenges	The Journal of Academic Librarianship	2017
Robin Rice	Research Data MANTRA: A Labour of Love	Journal of eScience Librarianship	2014
Ana Sesartic; Matthias Towe	Research Data Services at ETH-Bibliothek	IFLA Journal	2016
Shelley L. Knuth; Andrew M. Johnson; Thomas Hauser	Research Data Services at the University of Colorado Boulder	Bulletin of the Association for Information Science and Technology	2015
Carol Tenopir <i>et al.</i>	Research Data Services in Academic Libraries: Data Intensive Roles for the Future?	Journal of eScience Librarianship	2015
Carol Tenopir <i>et al.</i>	Research Data Services in European Academic Research Libraries	Liber Quarterly	2017
Carol Tenopir <i>et al.</i>	Research Data Services in European and North American Libraries: Current Offerings and Plans for the Future	Proceedings of the Association for Information Science and Technology	2016
Erin E. Kerby	Research data services in veterinary medicine libraries	Journal of the Medical Library Association	2016
Monica Lassi; Maria Johnsson; Koraljka Golub	Research data services: An exploration of requirements at two Swedish universities	IFLA Journal	2016
Anna Clements	Research information meets research data management ... in the library?	Insights	2013
Line Pouchard	Revisiting the Data Lifecycle with Big Data Curation	International Journal of Digital Curation	2015
Luis Martínez-Urbe; Paz Fernández	Servicios de datos: función estratégica de las bibliotecas del Siglo XXI	El profesional de la información	2015
Margaret E. Henderson; Teresa L. Knott	Starting a Research Data Management Program Based in a University Library	Medical Reference Services Quarterly	2015
Kevin B. Read <i>et al</i>	Starting the data conversation: informing data services at an academic health sciences library	Journal of the Medical Library Association	2015
Kristin Briney	Strategic Planning for Research Data Services	Bulletin of the Association for Information Science and	2016

		Technology	
Minglu Wang	Supporting the research process through expanded library data services	Program: electronic library and information systems	2013
Mary E. Piorun; Donna Kafel; Tracey Leger-Hornby	Teaching Research Data Management: An Undergraduate/Graduate Curriculum	Journal of eScience Librarianship	2012
Ryan Clement <i>et al.</i>	Team-based data management instruction at small liberal arts colleges	IFLA Journal	2017
Robert Fox	The art and science of data curation	OCLC Systems & Services: International digital library perspectives	2013
Shelley L. Knuth <i>et al.</i>	The Center for Research Data and Digital Scholarship at the University of Colorado-Boulder	Bulletin of the Association for Information Science and Technology	2017
Alex H. Poole	The conceptual landscape of digital curation	Journal of Documentation	2016
Li Si <i>et al.</i>	The cultivation of scientific data specialists: development of LIS education oriented to e-science service requirements	Library Hi Tech	2013
Abigail Goblen; Megan R. Sapp Nelson	The Data Engagement Opportunities Scaffold: Development and Implementation	Journal of eScience Librarianship	2018
Bryan Heidorn	The Emerging Role of Libraries in Data Curation and E-science	Journal of Library Administration	2011
Liz Lyon	The Informatics Transform: Re-Engineering Libraries for the Data Decade	International Journal of Digital Curation	2012
Tracy Gabridge	The Last Mile: Liaison Roles in Curating Science and Engineering Research Data	Research Library Issues	2009
Holly H. Yu	The role of academic libraries in research data service (RDS) provision: Opportunities and challenges	The Electronic Library	2017
Rebecca C. Reznik-Zellen; Jessica Adamick; Stephen McGinty	Tiers of Research Data Support Services	Journal of eScience Librarianship	2012
Elizabeth Berman	Transforming information literacy in the Sciences through the lens of e-Science	Communications in Information Literacy	2013
Krystyna K. Matusiak; Frank A. Sposito	Types of Research Data Management Services: An International Perspective	Proceedings of the Association for Information Science and Technology	2017
Amanda L. Whitmire; Michael Boock; Shan C. Sutton	Variability in academic research data management practices: Implications for data services development from a faculty survey	Program: electronic library and information systems	2015

Fonte: Elaborado pela autora.

APÊNDICE D – ROTEIRO DE ENTREVISTA

Caracterização dos serviços

Contextualize os serviços de apoio à gestão de dados de pesquisa na sua instituição. Se possível, aborde as seguintes questões:

- Origem e histórico;
- Precusores do atual serviço;
- Motivação para o oferecimento dos serviços;
- Existência de avaliação das necessidades dos pesquisadores na instituição e sua importância para o desenvolvimento dos serviços;
- Parcerias (internas e externas);
- Pessoas, unidades ou departamentos responsáveis pelo oferecimento dos serviços;
- Abrangência do serviço na instituição (toda a instituição ou apenas para algum projeto ou departamento);
- Infraestrutura tecnológica utilizada. Fatores que levaram à escolha dessa solução;
- Existência de política institucional para a gestão de dados de pesquisa;
- Financiamento para o desenvolvimento e funcionamento dos serviços;
- Tamanho e formação profissional da equipe.

Percepções

Comente sobre as dificuldades, barreiras, desafios, necessidades ou preocupações encontradas no oferecimento dos serviços em relação às questões mencionadas abaixo ou sobre qualquer outra questão que desejar.

- Capacidade da instituição e da biblioteca no oferecimento de serviços (pessoal, financeira, tecnológica ou técnica);
- Equipes (habilidades, competências, conhecimentos necessários e disposição em trabalhar no suporte à gestão de dados);
- Infraestrutura necessária;
- Financiamento dos serviços;
- Aceitação e apoio da instituição;
- Política institucional;
- Parcerias e colaborações;
- Contato com pesquisadores;
- Necessidades disciplinares específicas.

Serviços oferecidos e ações realizadas

Cite os serviços ou infraestruturas oferecidas, bem como ações realizadas para implantação e manutenção desses serviços (desenvolvimento de políticas, diretrizes, padrões, ferramentas *etc.*; grupos de discussão; divulgação; ou outra ação que considere importante).

APÊNDICE E – CATEGORIA 1 - INFRAESTRUTURA: TRECHOS

INF1-000 Gestão e curadoria – geral	[...] supporting collaborative technologies, or eResearch infrastructure development (KENNAN; CORRALL; AFZAL, 2014, p. 678, grifo nosso).
	A new system is currently under construction, which aims to support UQ researchers in managing their research data across the whole research lifecycle (YU; DEUBLE; MORGAN, 2017, p. 292).
INF1-001 Planejamento	Data management plan templates and tools (BERMAN, 2017c, p. 13).
INF1-002 Obtenção	[...] providing tools for dataset discovery [...] (SWANSON; RINEHART, 2016, p. 98).
INF1-004 Documentação de dados e metadados	Os bibliotecários podem oferecer soluções de armazenamento ou ferramentas que ajudam os pesquisadores a gerar os metadados durante a investigação (CORRÊA, 2016, p. 389, grifo nosso).
INF1-006 Armazenamento	The commitment of the Rectorate and Digital University Services of U.Porto to the development of the Institutional Repository has provided a solid ground for supporting an experimental data repository (RIBEIRO; FERNANDES, 2011, p. 16, grifo nosso).
	Preliminary projects in scientific disciplines generating and using mass data, like physics, meteorology or space science, have shown the importance of implementing a complementary infrastructure for the storage, management and provision of primary data . (OSSWALD, 2008, p. 518, grifo nosso).
INF1-011 Mineração, análise e visualização	[...] providing tools for data mining and visualization [...] (TENOPIR <i>et al.</i> , 2017, p. 27).
	Provision of advanced computing options (BERMAN, 2017c, p. 13).
INF2-012 Análise de redes sociais	[...] exploration and integration of digital scholarship tools and methods – such as social network analysis [...] (KNUTH <i>et al.</i> , 2017, p. 47, grifo nosso).
INF2-013 Assistência à produtividade e trabalho colaborativo	[...] include web-based tools for collaborative writing, electronic note-taking [...] (JOHNSON; BUTLER; JOHNSTON, 2012, p. 7, grifo nosso).
INF2-017 Gestão de pesquisas e comunicação científica	[...] research information system (CRIS) providing tools for managers and researchers to access all research-related institutional data from corporate systems such as human resources, student records, research grants and finance (CLEMENTS, 2013, p. 298, grifo nosso).
INF2-020 Políticas	The report, submitted to the SMT in July 2011, made a series of recommendations to improve data management practice. This included the creation of relevant institutional policies... (KNIGHT, 2015, p. 425-426, grifo nosso).
	A significant component of the EIDCSR Project was to develop and implement institutional policy on data management as part of a wider programme to promote research integrity [16]. This policy would also look to ensure compliance with funders' requirements. (WILSON <i>et al.</i> , 2010, não paginado, grifo nosso).
INF2-024 Tecnologias, infraestruturas e ferramentas	Research Electronic Data Capture (REDCap) is a rapidly evolving web tool for building and managing online surveys and databases (BARDYN <i>et al.</i> , 2018, p. 7, grifo nosso).
	[...] networking infrastructures allowing for storage and transport of data (PINFIELD; COX; SMITH, 2014, p. 28).

Fonte: Elaborado pela autora.

APÊNDICE F – CATEGORIA 2 - SUPORTE: TRECHOS

SUP1-000 Gestão e curadoria – geral	The Library will provide data services for acquisition, management, and preservation of massive amounts of data [...] (CLAIBOURN, 2015, p. 104, grifo nosso).
	[...] data management guidance during research (GOLDMAN; KAFEL; MARTIN, 2015, p. 2).
	[...] support the data curation [...] (GOLDMAN; KAFEL; MARTIN, 2015, p. 2).
	[...] assisting researchers with managing [...] (CORRALL; KENNAN; AFZAL, 2013, p. 655).
SUP1-001 Planejamento	Developing data management plans (CORRALL; KENNAN; AFZAL, 2013, p. 654).
	[...] assistance with finding funding sources [...] (KENNAN; CORRALL; AFZAL, 2014, p. 676).
	Guidance on writing data management plans (BERMAN, 2017c, p. 13).
SUP1-002 Obtenção	Data set purchasing (BERMAN, 2017c, p. 13).
	Finding relevant external data sets (CORRALL; KENNAN; AFZAL, 2013, p. 654).
	Libraries have to provide guidance on finding and assessing data [...] (KOLTAY, 2016, p. 98).
	Cornell University and Emory University Libraries began to help researchers locate and use existing datasets [...] (AKERS <i>et al.</i> , 2014, p. 182, grifo nosso).
SUP1-003 Organização e formato	Organizing Files (PERRIER <i>et al.</i> , 2018, p. 12).
	While libraries can help with giving structure and organization to how data are stored and labeled within repositories [...] (FREDERICK, 2016, p. 16, grifo nosso).
	[...] fornecendo informações sobre formatos de preservação [...] (CORRÊA, 2016, p. 389).
	[...] libraries provided information about how to properly organize data files (YOON; SCHULTZ, 2017, p. 926).
SUP1-004 Documentação de dados e metadados	Among others, these services comprise the creation of additional descriptive and administrative metadata for research data (VLAEMINCK, 2013, p. 71, grifo nosso).
	Guidance on how to use appropriate metadata Standards (BERMAN, 2017c, p. 13).
	[...] assistance with data documentation [...] (AKERS <i>et al.</i> , 2014, p. 175).
SUP1-005 Avaliação e seleção	[...] appraising datasets for curation and preservation (GOLDMAN; KAFEL; MARTIN, 2015, p. 6).
	Identifying data or datasets that could be candidates for repositories on or off campus (TENOPIR <i>et al.</i> , 2014, p. 87, grifo nosso).
	Assistance in selecting data to preserve for the long-term (BERMAN, 2017c, p. 13).
SUP1-006 Armazenamento	[...] and five (71%) libraries are identifying data or datasets for repositories, creating or transforming the metadata for these datasets and then ingesting the data or datasets into repositories (GOLDMAN; KAFEL; MARTIN, 2015, p. 6, grifo nosso).
	Librarians may provide some of that support, including identifying appropriate repositories , and advising on metadata and data management issues (MACMILLAN, 2014, p. 544, grifo nosso).
	Nesta fase, os bibliotecários podem ajudar os pesquisadores que não conhecem os requisitos para a preservação de dados, fornecendo informações sobre formatos de preservação, ou podem migrar dados diretamente (CORRÊA, 2016, p. 389, grifo nosso).
	Support for data deposit in an institutional repository (CORRALL; KENNAN; AFZAL, 2013, p. 654, grifo nosso).
	Assim, os bibliotecários podem prestar assistência aos investigadores, com o depósito de dados [...] (CORRÊA, 2016, p. 395, grifo nosso).
SUP1-007 Copyright, propriedade intelectual e licenciamento	Manage and handle personal, sensitive data and adhere to the protection requirements [...] License data (TRIPATHI; SHUKLA; SONKAR, 2017, p. 420, grifo nosso).
	[...] copyright and intellectual property assistance (CORRALL; KENNAN; AFZAL, 2013, p. 676).
	Libraries are providing consultation in many areas related to data management (Figure 2): all 16 (100%) responding libraries provide copyright services [...] (GOLDMAN; KAFEL; MARTIN, 2015, p. 6, grifo nosso).
SUP1-008 Questões éticas, legais e de consentimento	Some data may be sensitive, such as the locations of endangered species (e.g., rare orchids) that might be threatened if their locations were known. For data about humans, there may be privacy concerns. Some data might be restricted for national security and safety reasons. All of this means that the library must maintain proper controls on

	access (HEIDORN, 2011, p. 669, grifo nosso).
	Guidance on privacy and confidentiality (BERMAN, 2017b, p. 15).
SUP1-009 Publicação e citação	[...] promoting data citation [...] (KOLTAY, 2015, p. 402).
	[...] assistance with identifying journals in which to publish [...] (CORRALL; KENNAN; AFZAL, 2013, p. 677).
	[...] publication support [...] (MCLURE <i>et al.</i> , 2014, p. 144).
SUP1-010 Uso e reuso	[...] assigns DOIs to datasets [...] (AKERS <i>et al.</i> , 2014, p. 178).
	Acquiring unique identifiers for data sets (BERMAN, 2017c, p. 13).
	[...] assistance with open access issues and requirements (CORRALL; KENNAN; AFZAL, 2013, p. 676).
SUP1-011 Mineração, análise e visualização	[...] libraries as being central to the provisioning of data archiving, curation, and analysis of research data [...] (FOX, 2012, p. 173, grifo nosso).
	[...] assistance with data analysis and manipulation [...] (CORRALL; KENNAN; AFZAL, 2013, p. 677).
SUP2-013 Assistência à produtividade e trabalho colaborativo	Services with demand reported only by one library include assistance with writing for publication , the development of research project web services/ collaborative research environment [...] (CORRALL; KENNAN; AFZAL, 2013, p. 676-677, grifo nosso).
SUP2-014 Concepção e gestão de bases de dados	Other outcomes of this movement have been libraries offering consultations and workshops on data management planning, building databases for faculty projects [...] (VITALE, 2016, p. 39, grifo nosso).
	Database design and management (CHIWARE; BECKER, 2018, p. 7).
SUP2-015 Desenvolvimento de diretrizes, padrões, serviços e infraestruturas	[...] advise and help develop publishing standards and systems (GOLDb, 2007, não paginado).
SUP2-017 Gestão de pesquisas e comunicação científica	[...] providing guidance for developing workflows [...] (READ <i>et al.</i> , 2015, p. 131).
	An even more varied list of services were listed as having growing demand by two responding libraries: services related to data publication, grant application support, research information management [...] (KENNAN; CORRALL; AFZAL, 2014, p. 676, grifo nosso).
	[...] and offer guidance in scholarly communication, including authors' rights, open access and publishing in the institutional or other repositories (KNUTH <i>et al.</i> , 2017, p. 47, grifo nosso).
SUP2-019 Participação em projeto ou processo de pesquisa	[...] directly participating with researchers as a team member on a project (GOLDMAN; KAFEL; MARTIN, 2015, p. 6, grifo nosso).
	[...] participating with researchers on a project (as a team member) (TENOPIR <i>et al.</i> , 2014, p. 87, grifo nosso).
SUP2-020 Políticas	Assisting in developing policies about data management and preservation (BROWN; WOLSKI; RICHARDSON, 2015, p. 226).
SUP2-021 Rastreamento de impacto	[...] assessing the impact of research data (with research support offices) (KOLTAY, 2017, p. 7, grifo nosso).
	[...] collect impact evidence directly relating to research datasets and their reuse (LYON, 2012, p. 130, grifo nosso).
	A review of RDM websites shows the services offered include assistance with data collection , [...] collecting metrics to show the impact of shared data [...] (HENDERSON; KNOTT, 2015, sem paginação, grifo nosso).
SUP2-022 Reprodutibilidade da pesquisa	Whether libraries can claim that their existing knowledge prepares them to provide reproducibility services has not been fully explored. While it is true that libraries are well positioned as neutral in the academic landscape to provide this support [...] (VITALE, 2016, p. 41, grifo nosso).
SUP2-023 Serviços	When the libraries provided any service or information through their webpages , they also relied on or referred to other internal or external sources . [...] The most common type of internal link was a link to institutional repositories and resources [...] (YOON;

	SCHULTZ, 2017, p. 928, grifo nosso).
SUP2-024 Tecnologias, infraestruturas e ferramentas	[...] include web-based tools for collaborative writing, electronic note-taking, sharing citations, and assistance with mobile devices (JOHNSON; BUTLER; JOHNSTON, 2012, p. 7, grifo nosso).
	Consult on survey instruments (WANG, 2013, p. 289).
	[...] assistance with technology, infrastructure and tools [...] (CORRALL; KENNAN; AFZAL, 2013, p. 672).
	Statistical software support (WANG, 2013 p. 289).

Fonte: Elaborado pela autora.

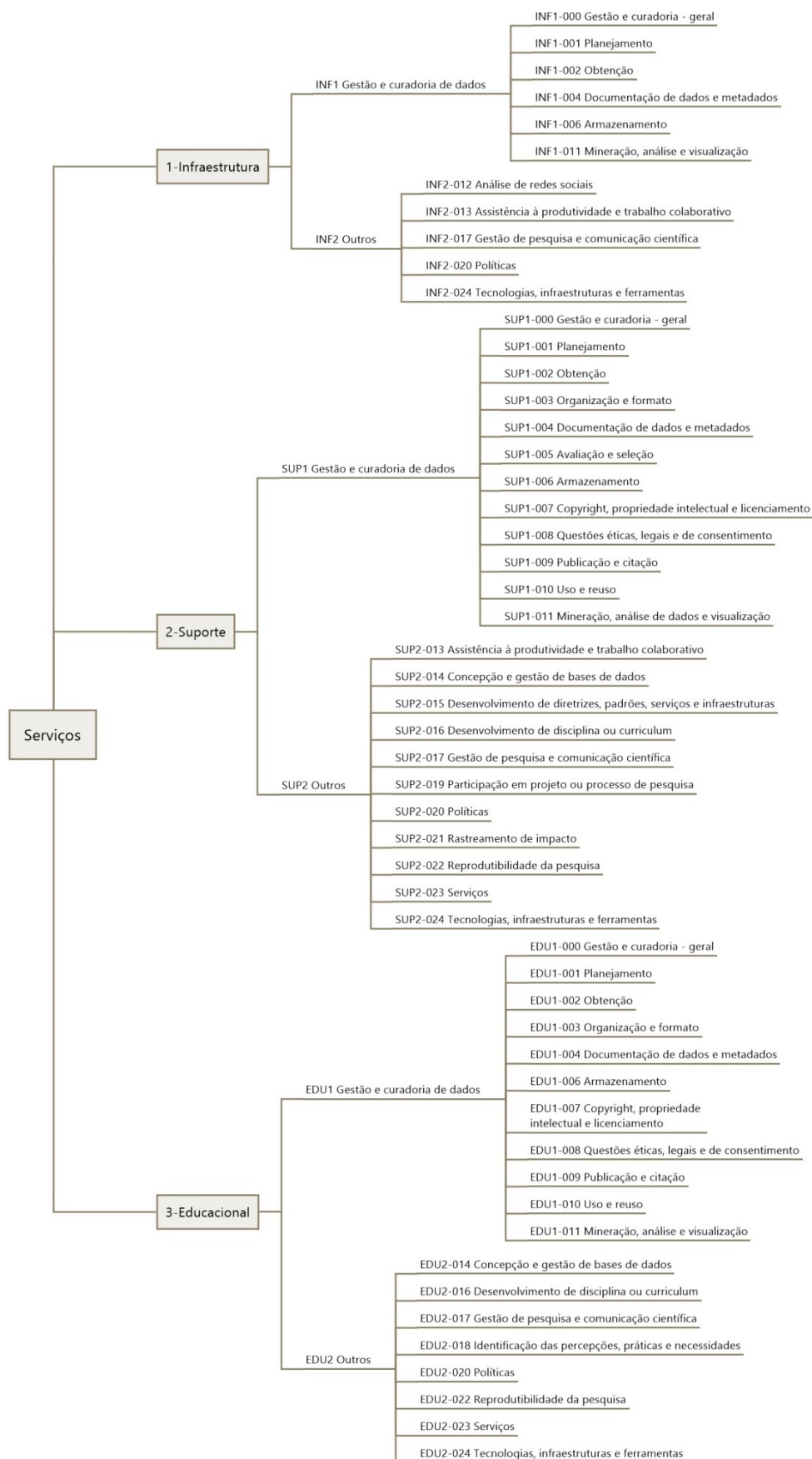
APÊNDICE G – CATEGORIA 3 - EDUCACIONAL: TRECHOS

EDU1-000 Gestão e curadoria - geral	Training is arguably the aspect of work that most clearly necessitates strong intra institutional coordination (WILSON <i>et al.</i> , 2011, p. 284).
	[...] offering workshops and classes on research data management [...] (GOLDMAN; KAFEL; MARTIN, 2015, p. 5).
EDU1-001 Planejamento	Data management plan workshops (BERMAN, 2017c, p. 13).
	Data management plan templates (BERMAN, 2017c, p. 13).
	[...] training classes on writing a DMP as part of the research grant application process [...] (YU; DEUBLE; MORGAN, 2017, p. 290).
EDU1-002 Obtenção	A key area where librarians can take a leadership role is instruction in the use of data resources. Librarians at a number of institutions have developed workshops in data location [...] (MACMILLAN, 2014, p. 547, grifo nosso).
	[...] the importance of providing training to researchers in accessing [...] (BROWN; WOLSKI; RICHARDSON, 2015, p. 227).
EDU1-003 Organização e formato	A workshop on types, formats, and stages of data [...] (PETERS; VAUGHN, 2014, p. 95).
	[...] training researchers on how to better organize [...] (READ <i>et al.</i> , 2015, p. 131).
EDU1-004 Documentação de dados e metadados	[...] providing information, consulting, training or active involvement in: data management planning, data management guidance during research, research documentation and metadata [...] (GOLDMAN; KAFEL; MARTIN, 2015, p. 2, grifo nosso).
	These services demonstrate the role of librarians and their impact on many areas of research, including the role as advocates of scholarly communication related to e-science, open access policies, open data, and the benefits of metadata standards (JOHNSON; BUTLER; JOHNSTON, 2012, p. 7, grifo nosso).
EDU1-006 Armazenamento	[...] librarians can educate and advocate for more consistent data sharing practices and standards [...] (NICHOLSON; BENNETT, 2011, p. 514, grifo nosso).
	This session [training] is now organised by the Graduate School, covering issues related to research activities at various stages of the research lifecycle. These issues include data storage options [...] (YU; DEUBLE; MORGAN, 2017, p. 293, grifo nosso).
EDU1-007 Copyright, propriedade intelectual e licenciamento	Libraries today play a vital role in advocating for intellectual property policies for scholarly texts that serve the advancement of ideas and knowledge. They can also play a role advocating for the documentation of rights and intellectual property in relation to data [...] (GOLD, 2007, online, grifo nosso).
EDU1-008 Questões éticas, legais e de consentimento	A workshop on types, formats, and stages of data (63%) follows closely behind, and there is interest in archiving & preservation (48%) and metadata (48%), but less interest in data sharing & re-use policies (38%) and legal and ethical considerations [...] (PETERS; VAUGHN, 2014, p. 95, grifo nosso).
	This [RDM training] includes an understanding of the ethics approval and informed consent process for human studies [...] (YU; DEUBLE; MORGAN, 2017, p. 291).
EDU1-009 Publicação e citação de dados	Some programs also briefly touched upon topics relevant to data management (such as publishing your data) during data-related classes or workshops (YOON; SCHULTZ, 2017, p. 928).
EDU1-010 Uso e reuso	Librarians are educating researchers about open access [...] (GOLDMAN; KAFEL; MARTIN, 2015, p. 2).
	In addition to consultation services, the JHUDMS also provides campus-wide training in best practices for RDM, with more specialized training sessions (e.g., managing data with personal identifiers [...] (AKERS <i>et al.</i> , 2014, p. 176, grifo nosso).
EDU1-011 Mineração, análise e visualização	[...] offering workshops and individual consultations on the use of statistical software, advanced data analytic techniques [...] (CLAIBOURN, 2015, p. 100).
	[...] introductory workshops on data processing and analysis Technologies [...] (YU; DEUBLE; MORGAN, 2017, p. 293).
EDU2-014 Concepção e gestão de bases de dados	Other outcomes of this movement have been libraries offering consultations and workshops on data management planning, building databases for faculty projects and actively managing data, among others (VITALE, 2016, p. 39, grifo nosso).

EDU2-016 Desenvolvimento de disciplina ou currículo	Faculty, researchers, graduate students and <i>librarians</i> from both institutions partnered on an IMLS planning grant to develop frameworks for an openly accessible data management curriculum for undergraduate and graduate students and researchers from a broad range of science, health sciences, and technology fields (KAFEL; CREAMER; MARTIN, 2014, p. 61, grifo nosso).
EDU2-017 Gestão de pesquisas e comunicação científica	<i>Librarians are educating researchers about</i> open access [...] <i>scholarly communication which is “the system research and other scholarly writings are created, evaluated for quality, disseminated to the scholarly community and preserved for future use”</i> [...] (GOLDMAN; KAFEL; MARTIN, 2015, p. 2, grifo nosso).
EDU2-018 Identificação das percepções, práticas e necessidades	Understanding data curation needs (LEE; STVILIA, 2017, p. 5). [...] identify investigator needs (REED, 2015, p. 2).
EDU2-020 Políticas	Guidance and training on the policies and processes put in place around sensitive data are essential to ensure researchers understand the options available to them and requestors are aware of the actions they should undertake in order to access datasets (HIOM <i>et al.</i> , 2017, p. 34, grifo nosso).
EDU2-022 Reprodutibilidade e da pesquisa	Indeed, some libraries have partnered with the research office or research computing departments on campus to bring outside speakers to campus to discuss methods and tools for improving reproducibility (VITALE, 2016, p. 39-40, grifo nosso).
EDU2-023 Serviços	[...] training sessions at faculty and graduate students, as well as meeting with the Office of Research and Sponsored Programs to explain the new requirement and RDS’ available support services (RABOIN; REZNIK-ZELLEN; SALO, 2012, p. 136, grifo nosso).
EDU2-024 Tecnologias, infraestruturas e ferramentas	NVivo training (KENNAN; CORRALL; AFZAL, 2014, p. 686). [...] offering workshops and individual consultations on the use of statistical software [...] (CLAIBOURN, 2015, p. 100). [...] programming in Python [...] (CLAIBOURN, 2015, p. 103).

Fonte: Elaborado pela autora.

APÊNDICE H – TAXONOMIA DE SERVIÇOS



Fonte: Elaborado pela autora.

APÊNDICE I – TAXONOMIA DE SERVIÇOS: NÚMERO DE REFERÊNCIAS

	Níveis		167	2731
		1 - INFRAESTRUTURA	80	293
		2 - SUPORTE	140	1612
		3 - EDUCACIONAL	125	826

	1 - INFRAESTRUTURA		80	293
		INF1 Gestão e curadoria de dados	74	240
		INF1-000 Gestão e curadoria - geral	22	30
		INF1-001 Planejamento	9	12
		INF1-002 Obtenção	3	5
		INF1-004 Documentação de dados e metadados	1	1
		INF1-006 Armazenamento	62	181
		INF1-011 Mineração, análise e visualização	8	11
		INF2 Outros	22	53
		INF2-012 Análise de redes sociais	1	1
		INF2-013 Assistência à produtividade e trabalho colaborativo	6	13
		INF2-017 Gestão de pesquisa e comunicação científica	2	4
		INF2-020 Políticas	13	29
		INF2-024 Tecnologias, infraestruturas e ferramentas	5	6

	2 - SUPORTE		140	1612
		SUP1 Gestão e curadoria de dados	137	1490
		SUP1-000 Gestão e curadoria - geral	104	299
		SUP1-001 Planejamento	91	221
		SUP1-002 Obtenção	61	145
		SUP1-003 Organização e formato	36	62
		SUP1-004 Documentação de dados e metadados	67	158
		SUP1-005 Avaliação e seleção	24	41
		SUP1-006 Armazenamento	91	267
		SUP1-007 Copyright, propriedade intelectual e licenciamento	25	44
		SUP1-008 Questões éticas, legais e de consentimento	17	23
		SUP1-009 Publicação e citação	40	73
		SUP1-010 Uso e reuso	48	93
		SUP1-011 Mineração, análise de dados e visualização	32	64
		SUP2 Outros	53	122
		SUP2-013 Assistência à produtividade e trabalho colaborativo	1	1
		SUP2-014 Concepção e gestão de bases de dados	2	2
		SUP2-015 Desenvolvimento de diretrizes, padrões, serviços e inf	5	7
		SUP2-016 Desenvolvimento de disciplina ou curriculum	3	3
		SUP2-017 Gestão de pesquisa e comunicação científica	5	5
		SUP2-019 Participação em projeto ou processo de pesquisa	17	24
		SUP2-020 Políticas	11	12
		SUP2-021 Rastreamento de impacto	13	19
		SUP2-022 Reprodutibilidade da pesquisa	1	3
		SUP2-023 Serviços	3	11
		SUP2-024 Tecnologias, infraestruturas e ferramentas	20	35

3 - EDUCACIONAL	125	826
EDU1 Gestão e curadoria de dados	120	753
EDU1-000 Gestão e curadoria - geral	109	533
EDU1-001 Planejamento	32	70
EDU1-002 Obtenção	7	7
EDU1-003 Organização e formato	6	6
EDU1-004 Documentação de dados e metadados	12	20
EDU1-006 Armazenamento	23	45
EDU1-007 Copyright, propriedade intelectual e licenciamento	2	3
EDU1-008 Questões éticas, legais e de consentimento	7	10
EDU1-009 Publicação e citação	5	9
EDU1-010 Uso e reuso	19	34
EDU1-011 Mineração, análise e visualização	3	16
EDU2 Outros	38	73
EDU2-014 Concepção e gestão de bases de dados	1	1
EDU2-016 Desenvolvimento de disciplina ou curriculum	6	10
EDU2-017 Gestão de pesquisa e comunicação científica	1	1
EDU2-018 Identificação das percepções, práticas e necessidades	4	14
EDU2-020 Políticas	3	3
EDU2-022 Reprodutibilidade da pesquisa	2	2
EDU2-023 Serviços	10	13
EDU2-024 Tecnologias, infraestruturas e ferramentas	16	29

Fonte: Elaborado pela autora.

APÊNDICE J – LISTA DE SERVIÇOS DA CATEGORIA 1 - INFRAESTRUTURA

1 INFRAESTRUTURA	
INF1 Gestão e curadoria de dados	
INF1-000 – Gestão e curadoria – geral	Fornecer ou desenvolver: • Infraestrutura de e-Research. • Infraestrutura interoperável para acesso, descoberta e compartilhamento de dados. • Plataforma para gestão e curadoria de dados de pesquisa (gerenciamento de metadados, submissão de dados ou conjunto de dados, armazenamento, arquivamento, preservação, gestão, publicação, compartilhamento, gerenciamento de usuários, controle de direitos, acesso, monitoramento de dados, análise de dados) durante todo o ciclo de vida dos dados, incluindo a gestão de dados ativos. • Ferramenta para auxílio na gestão de dados.
INF1-001 Planejamento	Fornecer ou desenvolver: • Plano de gestão de dados para a biblioteca ou para a instituição. • Ferramenta para a criação do PGD. • <i>Templates</i> para a criação do PGD. • Acesso à ferramenta para criação de PGD.
INF1-002 Obtenção	Fornecer ou desenvolver: • Catálogo de dados. • Ferramenta para descoberta e acesso aos conjuntos de dados. • Ferramenta para descoberta de ligações entre conjuntos de dados e documentos adjacentes, conjuntos de dados e pesquisadores e conjuntos de dados e seus financiadores.
INF1-004 Documentação de dados e metadados	Fornecer ou desenvolver: • Ferramenta para geração de metadados.
INF1-006 Armazenamento	Fornecer ou desenvolver: • Ferramenta para publicação de dados ou conjunto de dados. Repositórios de dados, específicos ou multidisciplinares – infraestrutura que suporte o arquivamento, compartilhamento, preservação e reuso de dados. • Serviços de repositórios como estatísticas e relatórios, preservação de dados e assistência com propriedade intelectual e direitos autorais. • Plataforma de troca e compartilhamento de dados. • Ferramentas ou serviços de <i>backup</i> de dados.
INF1-011 Mineração, análise e visualização	Fornecer: • Ferramenta para mineração de dados. • Ferramenta para visualização de dados. • Ferramenta para análise de dados. • Computação de alto desempenho (análise avançada, simulação, modelagem, inteligência artificial).
INF2 Outros	
INF2-012 Análise de redes sociais	Fornecer: • Ferramenta para análise de redes sociais.
INF2-013 Assistência à produtividade e trabalho colaborativo	Fornecer: • Ferramenta para anotações eletrônicas. • Ferramenta para compartilhamento de citações. • Plataforma para pesquisa colaborativa (compartilhamento e produção de conhecimento de forma colaborativa, escrita colaborativa).
INF2-017 Gestão de pesquisa e comunicação científica	Fornecer ou desenvolver: • Sistema para acesso a todos os dados institucionais relacionados à pesquisa - sistema de informações de pesquisa.

INF2-020 Políticas	Prover: • Políticas institucionais para a gestão de dados. • Políticas para acesso aos dados submetidos no repositório, licenciamento para reutilização de dados e período de embargo.
INF2-024 Tecnologias, infraestruturas e ferramentas	Fornecer: • Infraestruturas de rede que permitam o armazenamento e o transporte de dados. • Infraestrutura de computação avançada ou de alto desempenho. • Ferramenta para construção, administração e execução de <i>surveys</i>.

Fonte: Elaborado pela autora.

APÊNDICE K – LISTA DE SERVIÇOS DA CATEGORIA 2 - SUPORTE

2 SUPORTE	
SUP1 Gestão e curadoria de dados	
SUP1-000 Gestão e curadoria – geral	Fornecer suporte, orientação, assistência, aconselhamento, recomendações e consultoria para: • Gestão e curadoria de dados. • Preservação. • Gestão de dados ativos. • Problemas relacionados aos dados. • Melhores práticas para gestão de dados. • Administração de dados. • Manipulação e gestão de dados. • Serviço de referência em dados. • Coleções digitais. • Aspectos técnicos da curadoria digital. • eScience.
SUP1-001 Planejamento	Fornecer suporte, orientação, assistência, aconselhamento, recomendações e consultoria para: • Encontrar fontes de financiamento. • Pedido de financiamento (escrita e submissão). • Planejamento da gestão de dados antes de iniciar a pesquisa, incluindo as estratégias de preservação. • Escrita do plano de gestão de dados. • Método de pesquisa escolhido. • Estruturação da coleta tendo em vista o compartilhamento de dados. • Cumprimento dos requisitos para a gestão de dados das agências financiadoras e editores, incluindo requisitos de acesso aberto. • Cumprimento dos mandatos políticos.
SUP1-002 Obtenção	Fornecer suporte, orientação, assistência, aconselhamento, recomendações e consultoria para: • Desenvolvimento de protocolos para captura/coleta de dados ou criação de dados. • Captura/coleta ou criação de dados. • Identificação de fontes de dados. • Busca, localização e acesso a dados ou conjuntos de dados. • Compra/aquisição de conjuntos de dados. • Licenciamento ou assinatura de coleções de dados.
SUP1-003 Organização e formato	Fornecer suporte, orientação, assistência, aconselhamento, recomendações e consultoria para: • Limpeza de dados (ou pré-processamento de dados) identificação de inconsistências, redundâncias e anomalias nos dados ou conjunto de dados. • Identificação de formatos apropriados para preservação e acesso em longo prazo. • Formato de arquivos. • Conversão para formatos apropriados para a preservação. • Migração de dados. • Organização de arquivos de dados, incluindo convenções de nomenclatura. • Preparação de dados ou conjunto de dados para depósito. • Empacotamento de dados.

SUP1-004 Documentação de dados e metadados	Fornecer suporte, orientação, assistência, aconselhamento, recomendações e consultoria para: • Descrição de dados ou conjunto de dados. • Sistemas e padrões de descrição de dados que incentivem a interoperabilidade. • Criação e organização da documentação sobre os dados (dicionário de dados, descrição da metodologia <i>etc.</i>). • Criação e modificação de metadados. • Criação de padrões ou esquemas de metadados. • Identificação de padrões ou esquemas de metadados apropriados. • Padrões ou esquemas de metadados. • Aplicação de padrões ou esquemas de metadados.
SUP1-005 Avaliação e seleção	Fornecer suporte, orientação, assistência, aconselhamento, recomendações e consultoria para: • Identificação/Seleção/avaliação de dados ou conjunto de dados que devem ser preservados e incluídos no repositório. • Reavaliação de dados.
SUP1-006 Armazenamento	Fornecer suporte, orientação, assistência, aconselhamento, recomendações e consultoria para: • Armazenamento/arquivamento/depósito de dados ou conjunto de dados em repositório. • Implementação de repositório de dados. • Uso de repositórios de dados. • Depósito de dados em repositório internos e externos. • Identificação de repositórios apropriados para depositar dados. • Compartilhamento de dados. • Determinação do melhor repositório para um dado conjunto de dados. • Suporte técnico para repositórios de dados.
SUP1-007 Copyright, propriedade intelectual e licenciamento	Orientação, assistência, aconselhamento, recomendações e consultoria quanto: • Questões de <i>copyright</i> e propriedade intelectual. • Privacidade de dados. • Confidencialidade dos dados. • Licenciamento para acesso aberto. • Licenciamento de uso e reuso de dados. • Direitos de propriedade intelectual e material. • Propriedade dos resultados de pesquisa. • Direitos autorais e patentes. • Governança de dados (como questões de direitos autorais ou propriedade).
SUP1-008 Questões éticas, legais e de consentimento	Orientação, assistência, aconselhamento, recomendações e consultoria quanto: • Questões éticas e legais no que diz respeito à realização de pesquisas e à divulgação de dados. • Questões éticas e legais associadas aos conjuntos de dados. • Restrições de liberação e período de embargo. • Segurança e anonimização de dados de acordo com as políticas de conduta de pesquisa. • Políticas de dados institucionais (por exemplo, retenção, propriedade de propriedade intelectual). • Gestão e manipulação de dados pessoais e confidenciais. • Adoção de requisitos de proteção de dados. • Compartilhamento de materiais sensíveis. • Informações exigidas nos formulários de consentimento.

SUP1-009 Publicação e citação	Orientação, assistência, aconselhamento, recomendações e consultoria quanto: • Redação para publicação. • Publicação de dados – como publicar conjuntos de dados. • Identificação de periódicos nos quais publicar (data journals). • Citação de dados, conjunto de dados ou repositórios – como citar. • Padronização da citação de dados. • Promover a citação de dados.
SUP1-010 Uso e reuso	Orientação, assistência, aconselhamento, recomendações e consultoria quanto: • Requisitos de acesso aberto. • Adquirindo identificadores exclusivos para conjuntos de dados • Garantir acesso permanente aos dados • Garantir a descoberta e acessibilidade dos dados. • Discutir o acesso e uso de dados. • Assistência para obter um URL ou DOI para um conjunto de dados. • Garantir <i>links</i> apropriados de dados suplementares para artigos em repositórios institucionais.
SUP1-011 Mineração, análise de dados e visualização	Fornecer suporte, orientação, assistência, aconselhamento, recomendações e consultoria para: • Processamento de dados. • Simulação. • Modelagem de dados. • Extração de dados. • Manipulação de dados. • Apresentação de dados. • Visualização de dados. • Identificação de recursos para análise de dados. • Análise de dados. • Mineração de dados. • Métodos analíticos para dados. • Métodos estatísticos para dados.
SUP2 Outros	
SUP2-013 Assistência à produtividade e trabalho colaborativo	Fornecer suporte, orientação, assistência, aconselhamento, recomendações e consultoria para: • Desenvolvimento de um ambiente de pesquisa colaborativa.
SUP2-014 Concepção e gestão de bases de dados	Desenvolver: • Bases de dados para projetos do corpo docente.
SUP2-015 Desenvolvimento de diretrizes, padrões, serviços e infraestruturas	Participar, prestar consultoria, aconselhar, auxiliar no: • Desenvolvimento de repositório de dados. • Desenvolvimento de padrões de dados e metadados. • Desenvolvimento de normas e sistemas de publicação. • Desenvolvimento de plataforma para gestão de dados de pesquisa • Determinação de cotas de armazenamento para pesquisadores e outros depositantes. • Desenvolvimento de diretrizes sobre como as políticas serão implementadas e definindo atividades específicas, funções e responsabilidades.
SUP2-016 Desenvolvimento de disciplina ou <i>curriculum</i>	Participar, prestar consultoria, aconselhar, auxiliar na(s): • Discussões sobre como integrar o ensino de gestão de dados no currículo dos cursos de graduação e pós-graduação. • Integração da gestão de dados no currículo.

SUP2-017 Gestão de pesquisa e comunicação científica	Prestar assistência em: • Desenvolvimento de fluxos de trabalho e práticas de pesquisa. • Processos de comunicação científica.
SUP2-019 Participação em projeto ou processo de pesquisa	Participar ou fazer parceria para melhoria das práticas de gestão de dados nos: • Processos de pesquisa. • Projetos de pesquisa. • Grupos ou equipes de pesquisa.
SUP2-020 Políticas	Auxiliar, orientar: • Políticas institucionais relacionadas aos dados de pesquisa. • Conformidade com políticas, exigências legais e padrões éticos de pesquisa. • Desenvolvimento de política de gestão, preservação, compartilhamento e acessibilidade aos dados de pesquisa na instituição. • Desenvolvimento de políticas para liberação de dados e embargos. • Desenvolvimento de políticas em torno dos dados sensíveis.
SUP2-021 Rastreamento de impacto	Coletar, avaliar e demonstrar: • Impacto do compartilhamento e publicação de dados. • Impacto da reutilização de dados. • Estatísticas descritivas de uso de dados. Ajudar a: • Maximizar o impacto do pesquisador na <i>web</i>.
SUP2-022 Reprodutibilidade da pesquisa	Fornecer: • Orientações quanto à reprodutibilidade da pesquisa.
SUP2-023 Serviços	Fornecer orientação sobre os serviços voltados para a gestão de dados, internos ou externos (repositórios, <i>sites</i> , ou serviços).
SUP2-024 Tecnologias, infraestruturas e ferramentas	Oferecer suporte técnico, assistência, orientação ou consultoria para uso: • Geographic Information System (GIS). • Nvivo. • <i>Software</i> estatístico. • Ferramentas para manipulação de dados. • Sistemas de gestão de dados de pesquisa. • <i>Online survey</i>. • Dispositivos móveis. • Outras tecnologias, infraestruturas e ferramentas disponíveis.

Fonte: Elaborado pela autora.

APÊNDICE L – LISTA DE SERVIÇOS DA CATEGORIA 3 - EDUCACIONAL

3 EDUCACIONAL	
EDU1 Gestão e curadoria de dados	
EDU1-000 – Gestão e curadoria – geral	Oferecer treinamentos, <i>workshops</i>, cursos, seminários, conferências, aulas: • Data literacy. • Gestão de dados de pesquisa. • Curadoria de dados. • Melhores práticas de gestão de dados de pesquisa. Desenvolver guias, <i>templates</i>, materiais de apoio: • Gestão de dados de pesquisa. • Serviços para gestão de dados de pesquisa. Conscientização (<i>advocacy</i>): • Gestão de dados de pesquisa.
EDU1-001 Planejamento	Oferecer treinamentos, <i>workshops</i>, cursos, seminários, conferências, aulas: • Planejamento da gestão de dados. • Plano de gestão de dados - criação/escrita de planos. • Uso do DMPTool para criar planos de gestão de dados. • Requisitos dos financiadores em relação à gestão de dados e ao PGD. Desenvolver guias, modelos, materiais de apoio: • Plano de gestão de dados - criação/escrita de planos. Conscientização (<i>advocacy</i>): • Plano de gestão de dados.
EDU1-002 Obtenção	Oferecer treinamentos, <i>workshops</i>, cursos, seminários, conferências, aulas: • Planejamento da gestão de dados. • REDCap (ferramenta para captura de dados). • Digitalização de textos. • Localização e acesso aos dados ou conjunto de dados. Desenvolver guias, modelos, materiais de apoio: • Localização e acesso aos dados ou conjunto de dados. Conscientização (<i>advocacy</i>): • Fontes de dados reusáveis.
EDU1-003 Organização e formato	Oferecer treinamentos, <i>workshops</i>, cursos, seminários, conferências, aulas: • Tipos e formatos de dados. • Diferenças entre formatos de dados específicos. • Organização e nomeação de arquivos.
EDU1-004 Documentação de dados e metadados	Oferecer treinamentos, <i>workshops</i>, cursos, seminários, conferências, aulas: • Documentação de dados. • Descrição de dados. • Padrões de dados e metadados. • Implementação de padrões específicos de metadados. • Criação de metadados padronizados. Desenvolver guias, modelos, materiais de apoio: • Padrões de dados e metadados. Conscientização (<i>advocacy</i>): • Sistemas e padrões de descrição de dados com vistas à interoperabilidade.

EDU1-006 Armazenamento	Oferecer treinamentos, <i>workshops</i>, cursos, seminários, conferências, aulas: • Transferência de dados. • Arquivamento de dados. • Serviços de arquivamento. • Armazenamento de dados. • Opções para armazenamento de dados. • Compartilhamento de dados. • Compartilhamento de dados sensíveis. • Possibilidades e limitações do compartilhamento de dados. • Práticas e padrões de compartilhamento. • <i>Backup</i> de dados. • Segurança de dados. • Repositórios de dados. • Depósito de dados em um repositório. Desenvolver guias, tutoriais, modelos, materiais de apoio: • Repositórios de dados. • Serviços de repositório. • Compartilhamento de dados. • Requisitos de compartilhamento de dados. • Requisitos de retenção de dados ou conjunto de dados • Opções de armazenamento de dados. Conscientização (<i>advocacy</i>): • Requisitos de compartilhamento de dados.
EDU1-007 <i>Copyright</i>, propriedade intelectual e licenciamento	Oferecer treinamentos, <i>workshops</i>, cursos, seminários, conferências, aulas: • <i>Copyright</i> e propriedade intelectual. Conscientizar, educar: • <i>Copyright</i> e propriedade intelectual.
EDU1-008 Questões éticas, legais e de consentimento	Oferecer treinamentos, <i>workshops</i>, cursos, seminários, conferências, aulas: • Questões éticas e de consentimento. • Processo de aprovação de ética e consentimento informado para estudos em humanos. Desenvolver guias, tutoriais, modelos, materiais de apoio: • Questões éticas e de consentimento. Conscientizar, educar: • Privacidade de dados. • Requisitos de confidencialidade.
EDU1-009 Publicação e citação	Oferecer treinamentos, <i>workshops</i>, cursos, seminários, conferências, aulas: • Publicação de dados ou conjuntos de dados em repositórios. • Publicação de dados ou conjuntos de dados – como publicar. • Publicação de dados ou conjunto de dados – como a publicação beneficia as carreiras. • Citação de dados - como citar dados. Conscientizar, educar: • Citação de dados. • Publicação de data papers.

EDU1-010 Uso e reuso	Oferecer treinamentos, <i>workshops</i>, cursos, seminários, conferências, aulas: • Identificadores persistentes. • DOI. • Acesso aberto. • Dados abertos. • Uso de recursos de dados. • Reuso de dados. • Políticas para reuso de dados. • Replicação e disponibilidade de dados. • Desenvolver guias, tutoriais, modelos, materiais de apoio: • ORCID. Conscientizar, educar: • Ciência aberta. • Acesso aberto. • Políticas de acesso aberto. • Reuso de dados. • Identificadores pessoais.
EDU1-011 Mineração, análise e visualização	Oferecer treinamentos, <i>workshops</i>, cursos, seminários, conferências, aulas: • Processamento de dados. • Análise de dados qualitativos. • Técnicas analíticas avançadas. • Tecnologias de análise. • Preparação de dados para análise. • Visualização de dados.
SUP2 Outros	
EDU2-014 Concepção e gestão de bases de dados	Oferecer treinamentos, <i>workshops</i>, cursos, seminários, conferências, aulas: • Criação de bases de dados para projetos do corpo docente.
EDU2-016 Desenvolvimento de disciplina ou <i>curriculum</i>	Desenvolver disciplina ou currículo de gestão de dados para pesquisadores estudantes de graduação e pós-graduação.
EDU2-017 Gestão de pesquisa e comunicação científica	Conscientizar, educar: • Comunicação científica - a produção científica, sua criação, avaliação, disseminação e preservação para uso futuro.
EDU2-018 Identificação das percepções, práticas e necessidades	Conscientizar, educar: • Identificação e compreensão das necessidades do pesquisador relacionadas à gestão de dados.
EDU2-020 Políticas	Oferecer treinamentos, <i>workshops</i>, cursos, seminários, conferências, aulas: • Políticas institucionais, nacionais e internacionais. • Políticas e processos implementados em torno de dados sensíveis.
EDU2-022 Reprodutibilidade da pesquisa	Oferecer treinamentos, <i>workshops</i>, cursos, seminários, conferências, aulas: • Replicação e disponibilidade de dados. • Métodos e ferramentas para melhorar a reprodutibilidade da pesquisa.
EDU2-023 Serviços (equipe, pesquisadores)	Oferecer treinamentos, <i>workshops</i>, cursos, seminários, conferências, aulas: • Serviços de suporte disponíveis. • Serviços de dados. • Soluções de gestão de dados oferecidas na instituição. • Ferramentas e serviços para ajudar na gestão de dados de pesquisa.

EDU2-024 Tecnologias, infraestruturas e ferramentas	Oferecer treinamentos, <i>workshops</i>, cursos, seminários, conferências, aulas: • Ferramentas para análise de dados. • Ferramentas para gestão de dados. • Uso de <i>software</i> estatístico. • Análise de redes sociais. • Nvivo. • Geographic Information System (GIS). • Métodos estatísticos. • Introdução ao R. • Usando o LaTeX. • SQL. • Programação em <i>Python</i>. • Controle de versão com GIT. • <i>Bash scripting</i>. • Unix. • REDCap (captura de dados). • Uso de uma interface em particular. • EndNote. • Desenvolvimento de <i>web application</i>.
--	--

Fonte: Elaborado pela autora.