

**UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ
PROGRAMA DE PÓS-GRADUAÇÃO EM TECNOLOGIA E SOCIEDADE**

JANETE SALDANHA BACH ESTEVÃO

**LETRAMENTO INFORMACIONAL PARA REÚSO DE DADOS NAS CIÊNCIAS
SOCIAIS EM AMBIENTES VIRTUAIS DE PESQUISA: proposta de requisitos e
competências**

TESE

**CURITIBA
2019**

JANETE SALDANHA BACH ESTEVÃO

**LETRAMENTO INFORMACIONAL PARA REÚSO DE DADOS NAS CIÊNCIAS
SOCIAIS EM AMBIENTES VIRTUAIS DE PESQUISA: proposta de requisitos e
competências**

Tese apresentada como requisito parcial para a obtenção do Título de Doutor em Tecnologia e Sociedade pelo Programa de Pós-Graduação em Tecnologia e Sociedade da Universidade Tecnológica Federal do Paraná, Área de Concentração Tecnologia e Sociedade, Linha de Pesquisa Tecnologia e Desenvolvimento.

Orientadora: Profa. Dra. Faimara do Rocio Strauhs.

CURITIBA
2019

Dados Internacionais de Catalogação na Publicação

Estevão, Janete Saldanha Bach

Letramento informacional para reúso de dados nas Ciências Sociais em ambientes virtuais de pesquisa : proposta de requisitos e competências / Janete Saldanha Bach Estevão.-- 2019.

1 arquivo eletrônico (430 f.) : PDF ; 12 MB

Disponível via World Wide Web

Texto em português, com resumo em inglês

Tese (Doutorado) - Universidade Tecnológica Federal do Paraná. Programa de Pós-graduação em Tecnologia e Sociedade, Curitiba, 2019

Bibliografia: f. 308-336

1. Tecnologia - Teses. 2. Letramento informacional. 3. Teoria ator-rede. 4. Ciências sociais - Pesquisa. 5. Ciência da informação. 6. Comunicação na ciência. I. Strauhs, Faimara do Rocio, orient. II. Universidade Tecnológica Federal do Paraná - Programa de Pós-graduação em Tecnologia e Sociedade. IV. Título.

CDD: Ed. 23 – 600

Biblioteca Central da UTFPR, Câmpus Curitiba
Bibliotecária Anna T. R. Caruso CRB9/935

TERMO DE APROVAÇÃO DE TESE Nº 84

A Tese de Doutorado intitulada **LETRAMENTO INFORMACIONAL PARA REÚSO DE DADOS NAS CIÊNCIAS SOCIAIS EM AMBIENTES VIRTUAIS DE PESQUISA: proposta de requisitos e competências**, defendida em sessão pública pelo(a) candidato(a) **Janete Saldanha Bach Estevão**, no dia **22 de novembro de 2019**, foi julgada aprovada em sua forma final para obtenção do título de Doutor em Tecnologia e Sociedade, Área de Concentração – Tecnologia e Sociedade, Linha de Pesquisa – Tecnologia e Desenvolvimento, pelo Programa de Pós-Graduação em Tecnologia e Sociedade.

BANCA EXAMINADORA:

Prof^ª. Dr^ª. Sonia Elisa Caregnato - (UFRGS)
Prof. Dr. Décio Estevão do Nascimento - (UTFPR)
Prof^ª. Dr^ª. Marília Abrahão Amaral - (PPGTE/UTFPR)
Prof^ª. Dr^ª. Helena de Fátima Nunes Silva - (UFPR)
Prof^ª. Dr^ª. Faimara do Rocio Strauhs - (UTFPR) - Orientadora

A via original deste documento encontra-se arquivada na Secretaria do Programa, contendo a assinatura da Coordenação após a entrega da versão corrigida do trabalho.

Curitiba, 22 de novembro de 2019.

DEDICATÓRIA

Para todas e todos os estudantes da Pós-Graduação, que abraçam o ideal de obter uma formação, que esta lhes ajude a contribuir ainda mais com a sociedade, mesmo enfrentando tantos desafios na Academia para realizá-lo.

AGRADECIMENTOS

Escrever uma tese de doutorado não é uma tarefa que se faz sozinha, portanto, deixo meu reconhecimento e gratidão a todos e a todas que contribuíram para a viabilidade desta conquista.

À Universidade Tecnológica Federal do Paraná, ao PPGTE e ao corpo docente que, desde o mestrado, me proporcionaram uma formação superior pública, gratuita e de qualidade, abrindo a caixa preta da tecnologia e a da sua relação com a sociedade, transformando profundamente a minha percepção de mundo.

À orientadora Profa. Dra. Faimara do Rocio Strauhs, pelo compartilhamento do conhecimento e pelo engajamento neste trabalho.

À banca de qualificação e de defesa, Profa. Dra. Faimara do Rocio Strauhs, Profa. Dra. Sonia Elisa Caregnato, Profa. Dra. Helena de Fátima Nunes Silva, Profa. Dra. Marília Abrahão Amaral, Prof. Dr. Décio Estevão do Nascimento, pelo incentivo, pelas contribuições, pelos olhares mais aguçados que moldaram este trabalho e pelo reconhecimento de distinção e aprovação com louvor.

Ao Prof. Dr. Décio Estevão do Nascimento, pelo aprendizado da Teoria Ator-Rede, pelas oportunidades de contribuir com o grupo de pesquisa e pela sua acolhida.

Aos colegas do grupo de pesquisa Território – Redes, Políticas, Tecnologia e Desenvolvimento, pelas trocas e ensinamento mútuos, pelas ricas experiências compartilhadas.

À Universidade Federal do Paraná, pelo apoio e pelo incentivo para a minha dedicação exclusiva ao programa de estudos por dois anos.

À minha família, minha mãe Sonia Mara Saldanha Bach Passot, meu padrasto Pierre Jean Passot, meu esposo Juliano Estevão, meus filhos Benny Bach Estevão e Cristian Bach Estevão, pela compreensão, pelo apoio e pelas renúncias que esses quatro anos exigiram de todos nós.

Ao meu Senhor Jesus Cristo, por ter me conduzido nesta jornada em todos os momentos, dos mais felizes aos mais sombrios e solitários.

RESUMO

ESTEVIÃO, Janete Saldanha Bach. **Letramento informacional para reúso de dados nas Ciências Sociais em Ambientes virtuais de pesquisa**: proposta de requisitos e competências. 2019. 430 f. Tese (Doutorado em Tecnologia e Sociedade) – Programa de Pós-Graduação em Tecnologia e Sociedade, Universidade Tecnológica Federal do Paraná, Curitiba, 2019.

O ambiente da pesquisa científica vem sendo impactado pelo aumento da disponibilidade de dados em acesso aberto, impulsionado principalmente pela exigência do depósito dos dados a partir de leis governamentais, de políticas de órgãos de financiamento e de mandatos de publicações científicas. Neste cenário, emergem as dificuldades dos pesquisadores na gestão dos seus próprios dados e de terceiros, principalmente em relação às questões ligadas ao reúso dos mesmos. Cada campo da Ciência traz em seu bojo uma particularidade quanto aos tipos, aos volumes e às formas de produção de dados, inerente às suas próprias particularidades científicas. Porém, esta relação de disponibilidade *versus* gestão de dados é mais complexa nas Ciências Sociais, cujos dados são heterogêneos e variáveis. Põe-se a lume, a partir de fontes bibliográficas, a falta de letramento informacional em reúso de dados de pesquisa por parte dos pesquisadores, o que expõe as potencialidades inexploradas de reúso dos dados já disponíveis. E, pelas lentes da Teoria Ator-Rede (TAR), ressalta-se que os dados de pesquisa são em si um ator relevante no processo da comunicação científica, pois estes alteram a forma como os pesquisadores podem produzir análises, inferências e comparações a partir do reúso dos dados disponíveis nos ambientes virtuais de pesquisa, resultando em conhecimento novo. Este trabalho propõe, portanto, a partir da perspectiva da TAR, requisitos de um ambiente virtual para o letramento informacional em dados, com as competências mínimas para o reúso de dados em Ciências Sociais. A Pesquisa é de natureza aplicada, exploratório-descritiva combinada; parte-se de um levantamento bibliográfico, empregando-se técnicas de análise de conteúdo com apoio da ferramenta NVivo para análise e exploração do material em três instâncias de investigação. Na primeira instância, compara-se (1) os quadros conceituais (*frameworks*) de letramento informacional em dados selecionados, identificando quais são as características comuns propostas para o reúso de dados. Na segunda, reutiliza-se (2) dados secundários de pesquisa (*datasets*) sobre comportamento dos pesquisadores em compartilhar e em reutilizar dados em suas próprias pesquisas, evidenciando lacunas de letramento como barreiras ao reúso. Na terceira instância, identifica-se (3) nos Ambientes Virtuais de Aprendizagem de projetos selecionados, os recursos que ensinam os pesquisadores sobre reúso de dados. A partir do resultado da análise em triangulação dessas instâncias (*frameworks*, *datasets* e recursos de aprendizagem), este trabalho apresenta a proposta de 16 requisitos de um ambiente virtual para o letramento em reúso de dados contextualizadas nas disciplinas das Ciências Sociais e de 37 competências para os pesquisadores. Destacam-se como requisitos principais o acesso, a licença e o uso, e como competências essenciais as regras de acesso, os tipos de licenças e como usar os conjuntos de dados disponíveis. A validade da proposta foi testada por meio da verificação de aderência do conjunto de requisitos proposto com as melhores práticas em gestão de dados recomendadas pelo World Wide Web Consortium (W3C).

Palavras-chave: Ciência Aberta. Dados de Pesquisa – reúso. Letramento informacional em dados. Teoria Ator-Rede. Ambiente Virtual de Pesquisa.

ABSTRACT

ESTEVIÃO, Janete Saldanha Bach. **Information Literacy for data reuse in Social Sciences in Virtual Research Environments**: proposal of requirements and competences. 2019, 430 f. Thesis (PhD in Technology and Society) – Programa de Pós-Graduação em Tecnologia e Sociedade, Universidade Tecnológica Federal do Paraná, Curitiba, 2019.

The scientific research environment has been impacted by the increase availability of research data in open access, mainly due to the requirement of data deposit by government's laws, funding agencies' policies and scientific publication's mandates. In this scenario, the researchers face difficulties in managing their own data as well as third parties'; especially in data reuse related matters. Each field of Science has its own particularities of data types, volumes and forms of data production, which is inherent to each scientific particularities. However, the relation of availability versus data management is more complex in the Social Sciences, whose data are heterogeneous and variable. Therefore, this work seeks to evidence, from bibliographical sources, the lack of researcher's information data literacy, which exposes the untapped potential of data reuse already available. From the Actor-Network Theory (ANT) point of view, research data is itself a relevant actor in the process of scholarly communication, since they alter the way in which researchers can produce analyzes, inferences and comparisons, from data reuse available in virtual research environments, resulting in new knowledge. From an Actor-Network Theory (TAR) perspective, the research data are themselves a relevant actor in the process of scientific communication, since they alter the way researchers can produce analyzes, inferences and comparisons by reusing available data in virtual research environments, resulting in new knowledge. This work proposes, therefore, from an ANT perspective, the requirements of a virtual environment for informational data literacy, with the minimal necessary skills for data reuse in Social Sciences. The research is applied, exploratory-descriptive combined, based on a bibliographical background, and uses the NVivo analysis tool to support the content analysis techniques and material exploration into three research instances. For the first instance, it compares (1) the conceptual frameworks of information data literacy, identifying the common requirements proposed for data reuse within those frameworks. For the second instance, it reuses (2) secondary datasets about researchers' behavior in sharing and reusing data in their own research, highlighting literacy gaps as barriers to reuse. For the third instance, it identifies (3) in the Virtual Learning Environments of selected projects, what learning resources teach data literacy for reuse purposes. From the result of the triangulation analysis of these instances (frameworks, datasets and learning resources), this work presents a model of 16 requirements and 37 competences for research data reuse in the Social Sciences. Key requirements include access, licensing and use, and core competencies are access rules, license types and how to use available datasets. The validity test of the proposed model is performed by adherence checking with the best available data management recommendation practices by the World Wide Web Consortium (W3C).

Keywords: Open Science. Research Data – reuse. Information Data Literacy. Actor-Network Theory. Virtual Research Environment (VRE).

LISTA DE ILUSTRAÇÕES

Figura 1 - Taxonomia da Ciência Aberta – delimitação da pesquisa	26
Figura 2 - Frequência de conceitos-chaves – busca simples	41
Figura 3 - Combinação entre os descritores – quantidade recuperada e selecionada para a revisão.....	45
Figura 4 - Planejamento da pesquisa – macro etapas de pesquisa.....	53
Figura 5 - Estrutura da tese.....	56
Figura 6 - Disponibilidade dos dados de pesquisa como ponto de passagem obrigatório	73
Figura 7 - Percentual de alinhamento dos regulamentos em acesso aberto brasileiros ao H2020	84
Figura 8 - Nível de aderência de regulamento individual ao padrão de excelência H2020	85
Figura 9 - Etapas da gestão dos dados propostas pela inter-University Consortium for Political and Social	100
Figura 10 - Estrutura conceitual do ambiente de pesquisa	129
Figura 11 - Ambiente de pesquisa sob a perspectiva da Teoria Ator Rede.....	131
Figura 12 - Proposta conceitual da pesquisa	135
Figura 13 - Sequência das macro etapas da pesquisa	148
Figura 14 - Macro etapas de pesquisa do levantamento bibliográfico e da análise sistemática	150
Figura 15 - Desdobramentos do procedimento de revisão sistemática	151
Figura 16 - Relações de termos nas referências combinadas –WOS	161
Figura 17 - Inter-relação dos objetivos específicos e geral	184
Figura 18 - Etapas da análise de conteúdo	186
Figura 19 - As 50 palavras mais recorrentes	188
Figura 20 - Categorias criadas por palavras mais recorrentes	189
Figura 21 - Exemplo de variações de palavras mais recorrentes - <i>data</i>	190
Figura 22 - Palavras mais recorrentes – categorias de palavras mais recorrentes.....	191
Figura 23 - Relações semânticas dos termos selecionados- <i>citation</i>	196
Figura 24 - Exemplo de arquivo .CSV (1) transformado em arquivo em colunas (2)	205
Figura 25 - Estrutura do capítulo 6.....	210
Figura 26 - Infográfico das análises de Kim (2016 [<i>datasets</i>]).....	217
Figura 27 - Infográfico das análises de Kim (2017 [<i>datasets</i>]).....	221
Figura 28 - Infográfico das análises de Berghmans <i>et al.</i> (2017 [<i>datasets</i>]).....	226
Figura 29 - Recursos de Aprendizagem <i>Online</i> – codificação das categorias de análise.....	233
Figura 30 - Exemplos de problemas com codificação automática	234
Figura 31 - Problemas com codificação automática no <i>software</i> NVivo.....	235
Figura 32 - Conteúdos analisados nos recursos de aprendizagem <i>online</i>	236
Figura 33 - Conteúdos analisados nos dos recursos de aprendizagem <i>online</i>	237
Figura 34 - Conteúdos analisados nos dos recursos de aprendizagem <i>online</i>	239
Figura 35 - Critério de salvamento das buscas nos recursos de aprendizagem <i>online</i>	240
Figura 36 - Resultados salvos nos recursos de aprendizagem <i>online</i>	240
Figura 37 - Seleção do conteúdo para código manual dos recursos de aprendizagem <i>online</i>	241
Figura 38 - Exemplo de recorrências exportadas para análise	242

Figura 39 - Nuvem de palavras com 20 termos mais recorrentes nos recursos de aprendizagem	248
Figura 40 - Hierarquia da codificação das competências para reúso	252
Figura 41 - Fluxo da execução das atividades da pesquisa - 2017-2018.....	359
Figura 42 - Fluxo da execução das atividades de revisão sistemática da literatura.....	360

LISTA DE GRÁFICOS

Gráfico 1 - Levantamento bibliométrico – busca simples.....	158
Gráfico 2 - Descritores e combinação dos termos de pesquisa	160
Gráfico 3 - Seleção de artigos para revisão sistemática	172
Gráfico 4 - Publicações selecionadas - combinação C - por ano- 2012 a 2019	173
Gráfico 5 - Publicações selecionadas - combinação E – por ano- 2012 a 2019.....	175
Gráfico 6 - Publicações por base de dados - combinação C + E – 2012 a 2019	176
Gráfico 7 - Publicações por base de dados - combinação C + E + autores	176
Gráfico 8 - Tipos das referências utilizadas	177
Gráfico 9 - Referências utilizadas por ano	178
Gráfico 10 - Referências utilizadas- percentual por período.....	178
Gráfico 11 - Referências utilizadas por assunto	179
Gráfico 12 - Competências com mais de 100 recorrências - por recurso de aprendizagem...245	
Gráfico 13 - Recursos de Aprendizagem <i>Online</i> por Recorrência de Codificação -.....	246
Gráfico 14 - Recorrência das competências de acordo com os recursos de aprendizagem....	250
Gráfico 15 - Resultado da estratégia de busca do requisito <i>1 – Access</i>	253
Gráfico 16 - Resultado da estratégia de busca do requisito <i>16 - Using</i>	254
Gráfico 17 - Resultado da estratégia de busca do requisito <i>6 – Data Management</i>	255
Gráfico 18 - Resultado da estratégia de busca do requisito <i>08 - Licenses</i>	256
Gráfico 19 - Resultado da estratégia de busca do requisito <i>10 - Open Science</i>	257
Gráfico 20 - Aderência dos Requisitos às Recomendações W3C	295
Gráfico 21 - Publicações por ano - busca por autores - 2012 a 2019.....	342
Gráfico 22 - Publicações selecionadas - por autor - 2012 a 2019	342
Gráfico 23 - Autores e artigos mais citados - resumo	343

LISTA DE QUADROS

Quadro 1 - Comparativo das abordagens da Ciência Aberta.....	60
Quadro 2 - Comparativo dos conceitos das funções da comunicação científica.....	76
Quadro 3 - Apresentação comparativa dos <i>frameworks</i> das competências do letramento em dados.....	109
Quadro 4 - Recursos para o letramento em dados disponíveis em ambientes virtuais de aprendizagem.....	120
Quadro 5 - Resumo do alinhamento conceitual.....	133
Quadro 6 - Resumo de enquadramento metodológico	140
Quadro 7 - Matriz de consistência.....	144
Quadro 8 - Conceitos e estratégia de busca simples.....	154
Quadro 9 - Conceitos e estratégia de busca combinada	156
Quadro 10 - Teses e dissertações selecionadas	163
Quadro 11 - Amostra de artigos selecionados para a análise sistêmica	166
Quadro 12 - Combinação entre os descritores – quantidade recuperada.....	168
Quadro 13 - Protocolo geral de execução de pesquisa	181
Quadro 14 - Categorias de contexto	187
Quadro 15 - Amostra de palavras mais recorrentes na análise por grupo de <i>frameworks</i>	190
Quadro 16 - Exemplos de unidades de registro de <i>frameworks</i>	193
Quadro 17 - Exemplo de unidades de contextos de <i>frameworks</i>	194
Quadro 18 - Exemplo de agrupamento de categorias de análise de <i>frameworks</i>	195
Quadro 19 - OE2: Requisitos para o reúso de dados – análise dos quadros conceituais (<i>frameworks</i>) combinados	199
Quadro 20 - Requisitos aplicados nesta pesquisa para a análise dos dados secundários	200
Quadro 21 - OE3: Perspectivas e dificuldades dos pesquisadores no reúso de dados – análise dos dados secundários	228
Quadro 22 - Frequência de palavras nos Recursos de aprendizagem <i>online</i>	247
Quadro 23 - Requisitos identificados para reúso - recursos de aprendizagem.....	249
Quadro 24 - OE4: Competências para o reúso de dados – análise dos recursos de aprendizagem.....	251
Quadro 25 - OE5: Matriz dos requisitos comuns para reúso de dados – Triangulação	258
Quadro 26 - OG: Requisitos e competências para o letramento em reúso de dados.....	262
Quadro 27 - Indicador da estrutura dos requisitos e das competências essenciais.....	263
Quadro 28 - OE6: Aderência das recomendações das melhores práticas W3C.	293
Quadro 29 - OE6: Aderência dos requisitos às recomendações das melhores práticas W3C	294
Quadro 30 - Teses e dissertações selecionadas – resumo.....	337
Quadro 31 - Autores e artigos mais citados - Tenopir.....	344
Quadro 32 - Autores e artigos mais citados - Borgman	346
Quadro 33 - Autores e artigos mais citados - Koltay	348
Quadro 34 - <i>Framework</i> 01: competências do letramento em dados Carlson <i>et al.</i> (2011)....	361
Quadro 35 - <i>Framework</i> 02: competências do letramento em dados Calzada Prado e Marzal (2013)	364
Quadro 36 - <i>Framework</i> 03: objetivos de aprendizagem – Foster (2015).....	366

Quadro 37 - <i>Framework</i> 04: desenvolvimento profissional de pesquisadores em ensino superior – Vitae (2011).....	369
Quadro 38 - <i>Framework</i> 05: <i>data information literacy project</i> – Carlson e Johnston (2015).371	
Quadro 39 - <i>Framework</i> 06: University of Minnesota <i>Data management course</i> – Jeffryes; Johnston (2013).	373
Quadro 40 - Levantamento de unidades de registro de <i>frameworks</i>	375
Quadro 41 - Levantamento de unidades de contextos de <i>frameworks</i>	377
Quadro 42 - Agrupamento das categorias de análise dos <i>frameworks</i>	383
Quadro 43 - Categorias de análise, unidades de registro e contexto para o reúso de dados segundo <i>frameworks</i> combinados.....	386
Quadro 44 - Recursos de Aprendizagem <i>Online</i> por Recorrência de Codificação	402
Quadro 45 - Competências de reúso por Recorrência de Codificação	403
Quadro 46 - Melhores práticas da W3C	405

LISTA DE TABELAS

Tabela 1 - Frequência de Conceitos-Chaves – Busca Combinada	42
Tabela 2 - Levantamento de Teses e Dissertações	43
Tabela 3 - Políticas de depósito e acesso das revistas científicas brasileiras	87
Tabela 4 - Frequência de conceitos-chaves – busca simples.....	158
Tabela 5 - Frequência de conceitos-chaves – busca combinada	159
Tabela 6 - Levantamento de teses e dissertações	162
Tabela 7 - Publicações científicas <i>stricto sensu</i> brasileiras sobre gestão de dados de pesquisa – 2012-2019.....	164
Tabela 8 - Seleção das referências para compor <i>corpus</i> de análise – busca por descritores ..	169
Tabela 9 - Seleção das referências para compor <i>corpus</i> de análise – busca por autores.....	170
Tabela 10 - Resumo da seleção final de artigos para revisão sistemática – busca por descritores + autores	171
Tabela 11 - Combinação C – <i>corpus</i> dinâmico de 2012 a 2019.....	173
Tabela 12 - Combinação E – <i>corpus</i> dinâmico de 2012 a 2019	174
Tabela 13 - Lista de palavras mais recorrentes na análise por <i>frameworks</i> individuais.....	192
Tabela 14 - Perguntas relacionadas ao tema metadados.....	214
Tabela 15 - Competências nos recursos de aprendizagem com menos de 100 recorrências..	238
Tabela 16 - Recorrências de todas as competências nos recursos de aprendizagem.....	243
Tabela 17 - Pesquisa por autores – registros recuperados	341
Tabela 18 - Pesquisa por autores – <i>corpus</i> dinâmico de 2012 a 2019.....	341
Tabela 19 - Autores e artigos mais citados - resumo.....	343
Tabela 20 - Etapas de seleção de artigos para revisão sistemática.....	349
Tabela 21 - Resultados dos conjuntos de dados analisados 1.....	390
Tabela 22 - Resultados dos conjuntos de dados analisados 2.....	393
Tabela 23 - Resultados dos conjuntos de dados analisados 3.....	394

LISTA DE SIGLAS E ABREVIATURAS

AADP	Acesso Aberto aos Dados de Pesquisa
ACRL	Association of College and Research Libraries
ALA	American Library Association
ANP	Agencia Nacional do Petroleo, Gas Natural e Biocombustiveis
ARL	Association of Research Libraries
AVA	Ambiente Virtual de Aprendizagem
BDEP	Brazil Exploration and Production Database
BDTD	Biblioteca Digital Brasileira de Teses e Dissertações
BPMN	Business Process Model and Notation 2.0
Capex	Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
CC	Creative Commons
CCPA	California Consumer Protection Act
CESSDA	Consortium of European Social Science Data Archives
CRADLE	Curating Research Assets and Data Using Lifecycle Education;
CRIA	Centro de Referência em Informação Ambiental
CTS	Ciência, Tecnologia e Sociedade
C&T	Ciência e Tecnologia
DataONE	Data Observation Network for Earth
DCC	Digital Curation Center
DDI	Data Documentation Initiative
DIADORIM	Diretório de Políticas de Acesso Aberto das Revistas Científicas Brasileiras
DIL	Data Information Literacy
DOAJ	Directory of Open Access Journals
DOI	Digital Object Identificador
DMP	Data Management Plan
DPIA	Data Privacy Impact Assessments
DPO	Data Protection Officer
EaD	Educação à distância
ERIC	European Research Infrastructure Consortium
FAIR	Findable, Accessible, Interoperable and Reusable
FAPESP	Fundação de Amparo à Pesquisa do Estado de São Paulo
FERPA	Family Educational Rights and Privacy Act of 1974
Foster	Facilitate Open Science Training For European Research
GDPR	General Data Protection Regulation
GLOBE	Global Collaboration Engine
H2020	Horizon 2020
HIPAA	Health Insurance Portability and Accountability Act of 1996
Ibict	Instituto Brasileiro de Informação em Ciência e Tecnologia
ICPSR	Inter-university Consortium for Political and Social Research
IEEE	Institute of Electrical and Electronics Engineers
IES	Instituições de Ensino Superior
IFLA	International Federation of Libraries Association
INPE	Instituto Nacional de Pesquisas Espaciais
IODP	International Ocean Discovery Program
LAI	Lei de Acesso à Informação
LGPD	Lei Geral de Proteção de Dados
LIBER	Ligue des Bibliothèques Européennes de Recherche

LIS	Library and Information Studies
MANTRA	Research Data and Management Training
MCTI	Ministério da Ciência, Tecnologia e Inovação
MEC	Ministério da Educação
METS	Metadata Encoding and Transmission Standard
NDA	Non Disclosure Agreements
NECDMC	New England Collaborative Data Management Curriculum
NIH	National Institutes of Health
NSF	National Science Foundation
OCDE	Organização para Cooperação e Desenvolvimento Econômico
OCLC	Online Computer Library Center
OE	Objetivo Específico
OG	Objetivo Geral
OECD	Organization for Economic Co-Operation and Development
ORFG	Open Research Funders Group
OST	Office of Science and Technology
PBAADP	Programa Brasileiro de Acesso Aberto aos Dados de Pesquisa
PGD	Plano de Gestão de Dados
PII	Personally Identifiable Information
PL	Projeto de Lei
PLOS	Public Library of Science
PLS	Projeto de Lei do Senado
PPGTE	Programa de Pós Graduação em Tecnologia e Sociedade
PPO	Ponto de Passagem Obrigatório
PREMIS	Preservation Metadata Maintenance Activity
RDA	Research Data Alliance
RDF	Research Description Framework
RDF	Researcher Development Framework
RDM	Research Data Management
RDS	Research Data Services
RI	Repositório Institucional
RNP	Rede Nacional de Ensino e Pesquisa
ROAR	Registry of Open Access Repositories
ROARMAP	Registry of Open Access Repository Mandates and Policies
RSL	Revisão Sistemática da Literatura
SciELO	Scientific Electronic Library Online
TAR	Teoria do Ator- Rede
TDR	Trusted Digital Repository
TIC	Tecnologia da Informação e da Comunicação
UFMG	Universidade Federal de Minas Gerais
UFPR	Universidade Federal do Paraná
W3C	World Wide Web Consortium
WorldClim	Global Climate Data

SUMÁRIO

1	INTRODUÇÃO	20
1.1	TEMA DE PESQUISA	20
1.2	DELIMITAÇÃO DE PESQUISA	25
1.3	PROBLEMÁTICA DE PESQUISA, PRESSUPOSTOS E TESE	28
1.4	OBJETIVOS DA PESQUISA	36
1.4.1	Objetivo Geral	36
1.4.2	Objetivos Específicos	36
1.5	JUSTIFICATIVAS TEÓRICO-PRÁTICAS E MOTIVAÇÃO PESSOAL	37
1.5.1	Justificativa Teórica	38
1.5.2	Justificativa Prática e Motivação Pessoal	45
1.6	ABORDAGEM METODOLÓGICA DA PESQUISA	51
1.7	EMBASAMENTO TEÓRICO	53
1.8	ESTRUTURA DA TESE	55
2	COMUNICAÇÃO CIENTÍFICA: da Ciência Aberta à Publicação, breve cenário	57
2.1	O PARADIGMA CIENTÍFICO DO SÉCULO XXI - A CIÊNCIA ABERTA E O eScience	57
2.2	COMUNICAÇÃO CIENTÍFICA COMO UM SISTEMA SOCIOTÉCNICO	61
2.3	DO PROCESSO DE COMUNICAÇÃO CIENTÍFICA À REDE DE ATORES	64
2.4	DESCRIÇÃO DOS ATORES NA REDE DE COMUNICAÇÃO CIENTÍFICA	66
2.4.1	Os Dados de Pesquisa como Ator Não Humano da Rede	67
2.4.2	As Interações dos Dados com os demais Atores da Rede de Comunicação Científica	68
2.4.3	Dados de Pesquisa como Ponto de Passagem Obrigatório – PPO	70
2.5	FUNÇÕES DA COMUNICAÇÃO CIENTÍFICA	75
2.6	ACESSO ABERTO – GÊNESE, DESENVOLVIMENTO E BENEFÍCIOS	77
2.7	ACESSO ABERTO NO BRASIL	80
2.7.1	Projetos de Lei em Nível Nacional sobre Acesso Aberto no Brasil	81
2.7.2	Políticas em Níveis Institucionais de Acesso Aberto no Brasil	83
2.7.3	Políticas de Acesso Aberto das Revistas Científicas Brasileiras	85
3	OS DADOS DE PESQUISA E O LETRAMENTO INFORMACIONAL	89
3.1	O QUE SÃO OS DADOS DE PESQUISA	89
3.2	O PROBLEMA DO EXCESSO DE DADOS E POSSÍVEIS SOLUÇÕES	92
3.3	LETRAMENTO EM DADOS – CONCEITOS E PRÁTICAS	94

3.4	LETRAMENTO EM DADOS COMO UM SERVIÇO DAS BIBLIOTECAS ACADÊMICAS	99
3.4.1	Características dos Research Data Services (RDSs)	100
3.4.2	Estudo dos Tipos de RDSs	103
3.5	<i>FRAMEWORKS</i> DE LETRAMENTO EM DADOS	108
3.5.1	<i>Frameworks</i> Baseados em Modelos Genéricos <i>versus</i> Especializados	113
3.6	RECURSOS EM AMBIENTES VIRTUAIS DE APRENDIZAGEM	116
4	ALINHAMENTO CONCEITUAL E PROPOSTA INICIAL	123
4.1	CONCEITO DE CAIXA PRETA E A COMUNICAÇÃO CIENTÍFICA	124
4.2	O AMBIENTE VIRTUAL DA REDE DE COMUNICAÇÃO CIENTÍFICA	125
4.3	PROPOSTA CONCEITUAL DA PESQUISA	132
5	METODOLOGIA DA PESQUISA	136
5.1	CLASSIFICAÇÃO FORMAL DA PESQUISA – SINTESE DESCRITIVA	136
5.2	MACROETAPAS DE PESQUISA	141
5.2.1	Planejamento da pesquisa	142
5.2.2	Desenvolvimento da Pesquisa	145
5.2.3	Resultados da Pesquisa	146
5.3	PROCEDIMENTOS DE LEVANTAMENTO BIBLIOGRÁFICO	149
5.3.1	Especificação do assunto - Etapa 1	152
5.3.2	Definição das estratégias de busca e recuperação da informação - Etapa 2	152
5.4	RESULTADOS DO LEVANTAMENTO BIBLIOGRÁFICO	157
5.5	PROCEDIMENTOS DE REVISÃO SISTEMÁTICA DA LITERATURA	164
5.5.1	Seleção dos documentos - Etapa 3	165
5.5.2	Extração e tratamento das informações para redação do trabalho – Etapas 4, 5 e 6	167
5.6	RESULTADO DA REVISÃO SISTEMÁTICA	167
5.6.1	Etapas da Revisão Sistemática de Literatura (RSL)	170
5.6.2	Consolidação da Revisão Sistemática: <i>Corpus</i> teórico utilizado	177
5.7	METODOLOGIA DE ANÁLISE DE DADOS QUALITATIVOS - PROTOCOLO DE EXECUÇÃO	180
5.8	PROCEDIMENTOS DE COLETA E DE ANÁLISE DE DADOS SECUNDÁRIOS	184
5.8.1	As técnicas de análise – análise de conteúdo e de dados	185
5.9	PROCEDIMENTOS DE ANÁLISE DE <i>FRAMEWORKS</i> DE LETRAMENTO EM DADOS: requisitos de reuso	188
5.9.1	Análises dos <i>Frameworks</i> sobre Letramento em Dados	196

5.10	PROCEDIMENTOS DE ANÁLISES DE DADOS SECUNDÁRIOS – CONJUNTOS DE DADOS	200
5.11	PROCEDIMENTOS DAS ANÁLISES DOS RECURSOS DE APRENDIZAGEM <i>ONLINE</i>	207
6	APRESENTAÇÃO E ANÁLISES DE DADOS	211
6.1	RESULTADOS DAS ANÁLISES DOS CONJUNTOS DE DADOS SOBRE REÚSO DE DADOS.	211
6.1.1	Resultados das Análises dos <i>Datasets</i> – conjunto de dados de Kim (2016 [<i>datasets</i>])	211
6.1.2	Resultados das Análises dos <i>Datasets</i> – conjunto de dados Kim (2017 [<i>datasets</i>])	218
6.1.3	Resultados das Análises dos <i>Datasets</i> – conjunto de dados Berghmans <i>et al.</i> (2017 [<i>datasets</i>])	222
6.1.4	Consolidação dos Resultados das Análises dos <i>Datasets</i>	227
6.2	RESULTADOS DAS ANÁLISES DOS RECURSOS DE APRENDIZAGEM <i>ONLINE</i>	233
6.2.1	Apresentação dos Resultados dos Recursos de Aprendizagem - Resultados Gerais	233
6.2.2	Resultados dos Recursos de Aprendizagem – segmentação requisitos para reúso	248
6.3	MATRIZES DAS ANÁLISES DE CONTEÚDO	257
6.4	PROPOSTA DE UM CONJUNTO DE REQUISITOS PARA LETRAMENTO EM DADOS	260
6.5	VALIDAÇÃO DO GRAU DE ADERÊNCIA ÀS RECOMENDAÇÕES W3C	291
7	CONSIDERAÇÕES FINAIS	297
7.1	RESPOSTA À QUESTÃO DE PESQUISA E AOS OBJETIVOS	297
7.2	LIMITAÇÕES DA PESQUISA	303
7.3	SUGESTÕES PARA TRABALHOS FUTUROS E ÚLTIMAS CONSIDERAÇÕES	305
	REFERÊNCIAS	308
	APÊNDICE A – Referências análise sistemática – teses e dissertações	337
	APÊNDICE B – Autores e artigos mais citados	341
	APÊNDICE C – Seleção de artigos para revisão sistemática – 2018	349
	APÊNDICE D – Plano de gestão de dados (<i>Data Management Plan –DMP</i>)	350
	APÊNDICE E – Macro etapas de pesquisa - fluxo da execução das atividades	359
	APÊNDICE F – Revisão sistemática - fluxo da execução das atividades	360
	APÊNDICE G – <i>Frameworks</i> das competências do letramento em dados	361
	APÊNDICE H – Unidades de registro de <i>frameworks</i>	375
	APÊNDICE I – Unidades de contexto de <i>frameworks</i>	377
	APÊNDICE J – Agrupamento e formação das categorias de análise	383

APÊNDICE K – Resultados dos <i>frameworks</i> combinados	386
APÊNDICE L – Resultados do reúso de dados de Kim (2016 [<i>datasets</i>])	390
APÊNDICE M – Resultados do reúso de dados de Kim (2017 [<i>datasets</i>])	393
APÊNDICE N – Resultados do reúso de dados de Berghmans <i>et al.</i>, 2017 [<i>datasets</i>])	394
APÊNDICE O – Estratégia de Busca Textual – Recursos de Aprendizagem Online	399
APÊNDICE P – Recorrência de todos os requisitos em cada recurso de aprendizagem <i>online</i>	402
APÊNDICE Q – Recorrência das competências de reúso em cada recurso de aprendizagem <i>online</i>	403
APÊNDICE R – Data on the web best practices – W3C	405
ANEXO A – Instrumento de coleta de dados 1	407
ANEXO B – Instrumento de coleta de dados 2	411
ANEXO C – Instrumento de coleta de dados 3	415
ÍNDICE ONOMÁSTICO	424

1 INTRODUÇÃO

Este capítulo introdutório apresenta o tema de pesquisa que é situado no âmbito da Ciência Aberta. A delimitação do tema e a problemática tratam da competência informacional necessária para o reuso dos dados de pesquisa gerados nas Ciências Sociais, portanto, do letramento necessário para este. Apresenta ainda os objetivos e a proposta conceitual que permitem sua aplicação em ambientes de pesquisa e de aprendizagem.

Em seguida são apresentadas as justificativas sob aspectos teóricos e práticos da produção e da comunicação científicas, as escolhas e os procedimentos metodológicos para atingir os objetivos, detalhando as etapas em que a pesquisa foi construída. O embasamento teórico, abrangendo as teorias de base dos conceitos tratados está descrito na sequência. Por fim é apresentada a estrutura deste estudo.

1.1 TEMA DE PESQUISA

A Ciência Aberta representa uma nova abordagem para o processo científico com base no trabalho cooperativo e em novas formas de difusão de conhecimentos por meio de tecnologias digitais e de novas ferramentas colaborativas (EUROPEAN COMMISSION, 2016c). A Ciência Aberta é frequentemente tratada como um termo abrangente que envolve vários movimentos tendentes a suprimir as barreiras para o compartilhamento de qualquer tipo de resultados, de recursos ou de dados de pesquisa, métodos ou ferramentas, em qualquer fase do processo de investigação científica (ORGANIZATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT, 2015).

A partir da perspectiva de um conceito mais amplo da Ciência Aberta, as discussões sobre a disponibilidade dos dados de pesquisa se intensificaram (SAYÃO; SALES, 2015b). Igualmente, são muito diversificadas as abordagens dos estudos sobre o acesso aberto aos dados de pesquisa discutindo-se aspectos como:

- governança (KOLTAY; HORSTMANN; WITT, 2016);
- interoperabilidade (AMORIM *et al.*, 2015);
- políticas públicas de acesso aberto (EUROPEAN COMMISSION, 2016a; NATIONAL SCIENCE BOARD, 2005; NATIONAL SCIENCE FOUNDATION, 2007);
- preservação digital (SAYÃO, 2012; BORGMAN *et al.*, 2015);

- curadoria de dados (BORGMAN, 2012; POOLE, 2016);
- gestão e compartilhamento do conhecimento por meio do reúso de dados (CURTY, 2016a);
- melhor eficiência dos recursos públicos (KIMBROUGH, 2016);
- barreiras comportamentais e culturais em relação ao compartilhamento de dados (CURTY, 2016b);
- questões sobre direito autoral e propriedade intelectual (MONTEIRO, 2017);
- infraestrutura tecnológica e qualidade dos dados (SALES; SAYÃO, 2015a);
- custos de armazenamento e serviços de apoio à gestão de dados dos pesquisadores (TENOPIR; BIRCH; ALLARD, 2012; WHYTE *et al.*, 2014).

Na perspectiva de Sayão e Sales (2015b), o tema da Ciência Aberta está circundado em um ambiente complexo e de forma interdependente, mesclando o contexto de investigações e da comunicação científica, processo pelo qual o trabalho científico é criado, documentado, avaliado, disseminado e preservado (KIM, 2013). Segundo Kim (2013), mesmo que esse processo já venha sofrendo transformações desde a década de 2000, há um movimento maior para reformar o processo de comunicação científica por meio do suporte à promoção do acesso aberto aos dados em si.

A comunicação científica tem uma ampla abordagem, especialmente em duas vertentes: a publicação e a difusão dos resultados de pesquisa e o acesso ao material publicado (CREASER, 2011). A pesquisa em comunicação científica se intensificou nos anos 1960 e 1970, período em que diversos fluxos de informação de comunicação dos resultados de pesquisa foram propostos, especialmente com o trabalho seminal de Meadows (1974), assim como as obras de Lancaster (1977) e de Jordan (1973).

Meadows (1998, p. 5, tradução nossa) enfatizou que “a comunicação científica é o centro da pesquisa”, sendo essencial para que a pesquisa exista enquanto atividade social. Outros autores citados em Borgman (2007) tais como: Garvey (1979)¹; Lievrouw (1990)² e Paisley (1984)³ corroboram a visão que a comunicação científica é essencial para a pesquisa existir como um processo social. Para Lievrouw (1990, p. 59 *apud* BORGMAN, 2007), a comunicação científica é constituída de uma estrutura social que considera redes de

¹ GARVEY, W. D. **Communication**: The Essence of Science. New York: Pergamon, 1979.

² LIEVROUW, L. A. **Reconciling Structure and Process in the Study of Scholarly Communication**. In *Scholarly Communication and Bibliometrics*, ed. C. L. Borgman, p. 59–69. Newbury Park, CA: Sage, 1990.

³ PAISLEY, W. J. **Communication in the Communication Sciences**. In *Progress in the Communication Sciences*, ed. B. Dervin and M. Voigt, v.5, p. 1–44. Norwood, NJ: Ablex, 1984.

relacionamentos entre estudiosos de um campo, como os colégios invisíveis, as redes de conhecimento, as comunidades epistêmicas e as redes de cocitação. Outras entidades, no entanto, são englobadas como atores nesse processo: instituições de pesquisa, editores e bibliotecas, entre outros (JUBB; SHORLEY, 2013; LYMAN, 1999 *apud* BORGMAN, 2000; BORGMAN, 2007). De acordo com Borgman (2000; 2007), a comunicação científica é, portanto, um processo social constituído de diversos atores e dos relacionamentos entre estes, como: publicações acadêmicas, pesquisadores, instituições de pesquisa, editores, publicações, revisores, bibliotecas e bibliotecários.

Levando em conta a perspectiva da Teoria Ator-Rede (TAR), que considera os atores humanos e não humanos como actantes, ou seja, atores que influenciam as relações sociais a partir dos seus distintos tipos de agência, infere-se que os elementos propostos por Borgman (2000) podem ser considerados actantes, pois se configuram atores na medida que exercem influência no processo de comunicação científica, transformando tal processo. Exemplifica-se pela constatação de que as formas com as quais os pesquisadores divulgam suas pesquisas mudaram e se desenvolveram devido a uma variedade de aspectos sociais, de fatores econômicos e tecnológicos (KIM, 2013). O rápido crescimento da internet e do desenvolvimento da tecnologia da informação e da comunicação (TIC) reestruturaram, por exemplo, a dinâmica da publicação acadêmica, principalmente quanto ao modo de acesso e as fontes de informação (BORGMAN, 2000).

Constata-se, portanto, a agência da TIC como elemento não humano, interferindo no processo de comunicação científica, ou seja, a partir da evolução da tecnologia de informação (fator tecnológico), aliada ao impulso produzido pelo movimento de acesso aberto aos artigos de revistas (fatores social e econômico), propiciou-se igualmente um aumento significativo de fontes e de recursos virtuais de pesquisa (CREASER, 2011). De outra parte, com a constante intensidade na geração e no uso dos dados disponibilizados nessas comunicações, os arquivos digitais de dados têm tido um papel relevante, pois estão alterando a própria forma de construção da ciência (SAYÃO; SALES, 2012).

A partir dos anos 2000, as políticas que advogam pelo acesso aberto da comunicação da pesquisa científica financiada com recursos públicos têm impulsionado a difusão e o compartilhamento dos dados produzidos nos processos de investigação (CREASER, 2011). Nota-se nesta década, com movimento de forma mais intensa, que vários instrumentos foram estabelecidos para fomentar e também regular a gestão de dados de pesquisa, tais como a declaração de Berlin (BERLIN, 2003) e as políticas estabelecidas pelos órgãos de fomento à pesquisa (BERLIN, 2003; NATIONAL SCIENCE BOARD, 2005; NATIONAL SCIENCE

FOUNDATION, 2007; ORGANIZATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT, 2007).

No entanto, a força motora que tem, efetivamente, viabilizado a abertura dos dados de pesquisa têm sido as exigências de políticas institucionais das agências de fomento, em alguns casos de políticas nacionais como nos Estados Unidos e em países da Europa, políticas estas recorrentes entre as organizações mais expressivas da pesquisa científica no mundo (BORGMAN, 2000; CREASER, 2011; GRIGG, 2016).

Infere-se, entretanto, ainda que as exigências dos dados abertos de pesquisa sejam cada vez mais rígidas para a concessão de recursos públicos, que iniciativas e serviços de acesso aos dados e suporte ao reúso são tão necessários quanto a sua real disponibilidade. De outra forma, os repositórios permanecerão saturados de dados desconhecidos, inacessíveis e inutilizados, perdendo seu valor pela não utilização do seu potencial de geração de conhecimento a partir de novas abordagens.

Considera-se, portanto, que os dados de pesquisa são um actante, pois sua presença é marcada por uma série de alterações desdobradas e percebidas pela mudança que provoca na dinâmica de produção e de comunicação do conhecimento. Porém, a capacidade em reutilização destes dados é um requisito para atribuição de sentido e de valor aos dados. Essa capacidade requer do pesquisador o **letramento informacional em reúso de dados**. O letramento em dados, no entanto, já faz parte dos serviços que são de competência das bibliotecas universitárias (ASSOCIATION OF COLLEGE AND RESEARCH LIBRARIES, 2015).

Outro ator não humano no processo de comunicação científica que tem alterado sua forma de agência é a biblioteca acadêmica, que tem recebido a exigência de uma transformação no seu papel, para se adequar à realidade da pesquisa intensiva em dados, pois é uma entidade entrelaçada ao processo de pesquisa e está entre os principais recursos na estrutura de serviços necessários à gestão de dados (TENOPIR *et al.*, 2015a). Os serviços de gestão de dados consistem em serviços técnicos e informacionais que a biblioteca acadêmica oferece aos pesquisadores para gerenciar todo o ciclo de vida de seus dados (TENOPIR; BIRCH; ALLARD, 2012). A demanda por esses serviços já foi prevista pela Association of College and Research Libraries (ACRL) como uma das mais importantes tendências da atuação da biblioteca acadêmica no século XXI (ASSOCIATION OF COLLEGE AND RESEARCH LIBRARIES, 2014).

Aprofundando, o conceito de letramento informacional já está consolidado e é amplamente aceito como a habilidade para reconhecer a necessidade da informação, e de identificar, localizar, avaliar e usar informações para resolver um determinado problema

(AMERICAN LIBRARY ASSOCIATION, 1989). Johnson (2012) descreveu o letramento em dados com muito mais detalhes, conceituando-o como a capacidade de processar, classificar e filtrar grandes quantidades de informação, que exige saber como pesquisar, como filtrar e processar, produzir e sintetizar os dados. Entretanto, a terminologia no campo do letramento informacional em reuso de dados ainda não está padronizada por ser recente. A literatura passou a destacar o letramento em dados como parte do letramento informacional, porém, Koltay (2016) ressalta que muitos autores tratam os conceitos de forma autônoma. Ainda assim as fronteiras de informação em letramento em informação e letramento informacional em dados não são rígidas.

Na literatura o letramento em dados está referenciado com diversos termos, como: letramento em dados científicos (QIN; D'IGNAZIO, 2010), letramento em dados de pesquisa (SCHNEIDER, 2013) e letramento em informação de dados (CARLSON *et al.*, 2011), que conceitua este como a capacidade de coletar, usar e avaliar dados, de forma a "entender o que os dados significam, incluindo como ler gráficos e tabelas adequadamente, inferir conclusões corretas dos dados, e reconhecer quando os dados estão sendo usados de forma equivocada ou inadequada" (CARLSON; JOHNSTON, 2015, p. 15, tradução nossa).

No entanto, os pesquisadores de maneira geral não têm essas habilidades propostas pelo letramento em dados de forma satisfatoriamente desenvolvidas para corresponder às novas exigências do processo de produção científica, que incorpora os dados de pesquisa (CARLSON *et al.*, 2011; JEFFRYES; JOHNSTON, 2013; KOLTAY, 2017a; MAYBEE *et al.*, 2015).

Evidencia-se, portanto, uma lacuna de potencial estudo a partir da consideração da comunicação científica de um lado, como um processo social (GARVEY, 1979; LIEVROUW, 1990; PAISLEY, 1984; WHITLEY, 1984 *apud* BORGMAN, 2007), e de outro, a crescente complexidade do ambiente de pesquisa impulsionada pelo quarto paradigma, conduzido pela geração e pelo uso de grandes volumes de dados (GRAY, 2005). No centro está um importante ator, o cientista, que a partir do letramento em como gerenciar e utilizar os dados já disponíveis pode potencializar os resultados de suas pesquisas rumo à criação de conhecimento novo (BORGMAN, 2000; 2007; CURTY, 2016a, CURTY, 2016b; CURTY *et al.*, 2017; SAYÃO; SALES, 2014; TENOPIR, 2011, 2015;).

O foco deste relatório de tese, neste contexto, é justamente o acesso aberto aos dados de pesquisa já disponibilizados, prática que tem como base as concepções da Ciência Aberta. Para potencializar a aplicabilidade desses dados disponíveis é que se insere a necessidade do letramento em reuso de dados, por meio dos requisitos de um ambiente virtual que estimule as competências essenciais para este reuso, em especial neste estudo para a área das Ciências

Sociais.

1.2 DELIMITAÇÃO DE PESQUISA

Esta pesquisa limitou-se, diante de um tema amplo como o da Ciência Aberta, portanto, ao escopo do acesso aberto e ao estudo dos dados de pesquisa. Os dados de pesquisa, quando disponíveis e reutilizáveis, alteram a forma como os pesquisadores podem produzir análises, inferências e comparações, resultando em conhecimento científico novo. Este estudo está direcionado para a investigação no campo das Ciências Sociais, visto que há uma maior dificuldade de compartilhamento nesta área se comparado às Ciências Naturais, devido ao volume e a diversidade das fontes, além dos dados serem muito heterogêneos, de difícil replicação, pois são coletados em contextos de ambientes sociais distintos (BORGMAN, 2007).

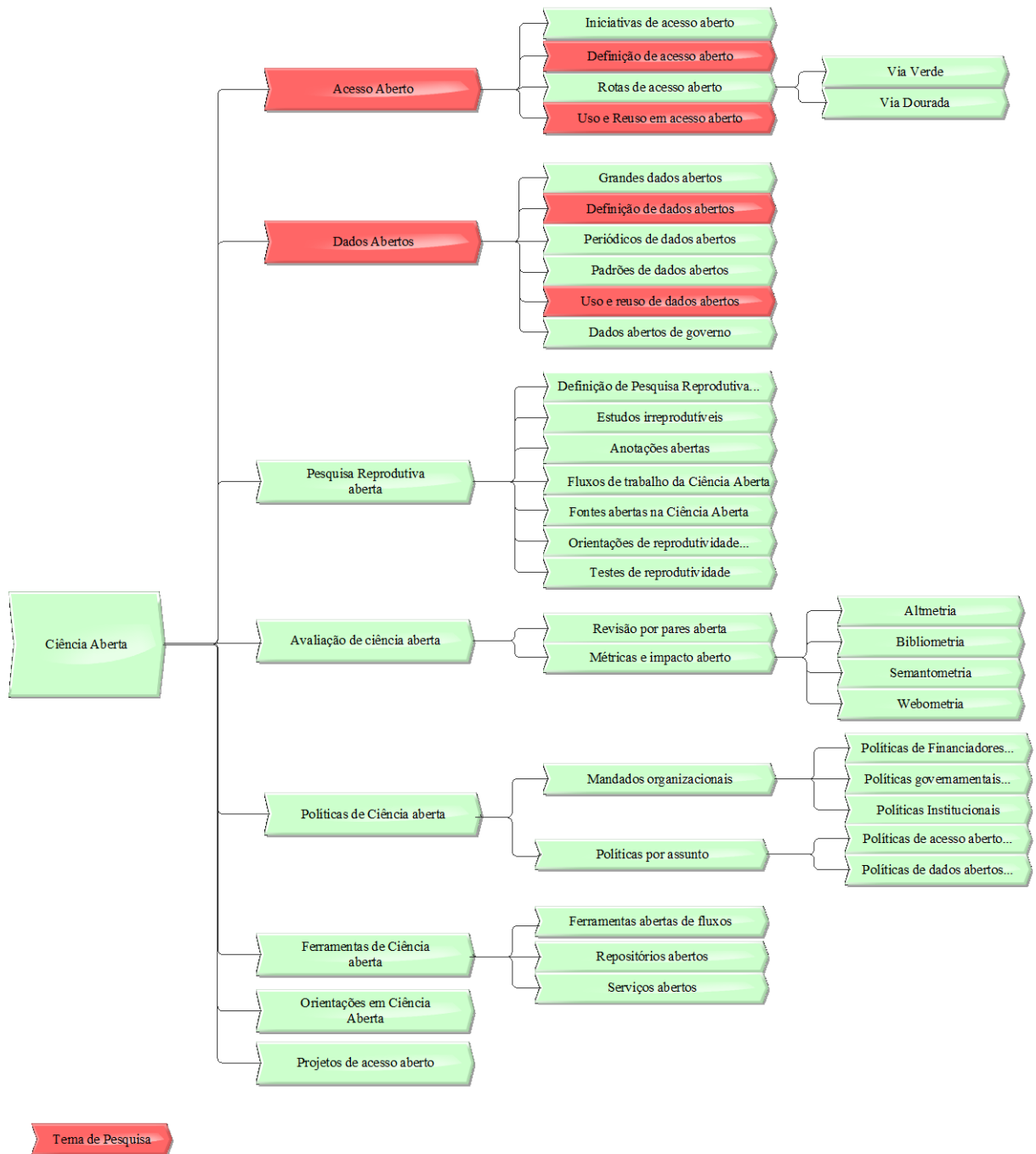
O domínio da Ciência Aberta está representado na taxonomia construída de acordo com a plataforma Facilitate Open Science Training For European Research (Foster). As taxonomias são ferramentas que podem ser usadas para classificar e organizar conteúdo em um dado domínio (PONTIKA *et al.*, 2015). Não foram localizadas taxonomias semelhantes em língua portuguesa, portanto, optou-se pela visualização da delimitação do tema deste trabalho a partir do uso da taxonomia para a Ciência Aberta de Pontika *et al.* (2015).

Na abordagem conceitual da Ciência Aberta, este estudo adota os preceitos das escolas democrática e pragmática de Fecher e Friesike (2014), que preconizam o acesso e a pesquisa colaborativa também para os dados abertos.

A delimitação deste trabalho está especificada, portanto: (i) no conceito-chave do acesso aberto e de dados abertos – uso e reúso dos dados, conforme destacados na Figura 1, e (ii) nos ambientes virtuais que viabilizem o letramento informacional para o reúso dos dados. Este ambiente considera os requisitos tecnológicos e os comportamentais dos pesquisadores, visando à geração de conhecimento necessário para reúso e extração de valor científico de conjuntos de dados disponíveis.

O conceito de letramento em dados adotado nesta pesquisa fundamenta-se no trabalho de Carlson *et al.* (2011), pois incorpora um conjunto de letramentos, como o de dados em si, o da estatística e o da informação enquanto **habilidades funcionais** em todo o ciclo de gestão de dados (CARLSON *et al.*, 2011). Em contrapartida, Koltay (2016) estabelece o letramento em dados como parte do letramento informacional, uma subdivisão deste, portanto, ainda que defenda limites não rígidos entre esses letramentos (KOLTAY, 2016).

Figura 1 - Taxonomia da Ciência Aberta – delimitação da pesquisa



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

No âmbito da Ciência Aberta, os dados abertos fazem parte dos elementos não humanos que se inter-relacionam com os humanos no processo de produção do conhecimento científico e na difusão dos resultados de pesquisa. Logo, considera-se a Teoria Ator-Rede como direcionador metodológico. Utilizar a perspectiva metodológica da Teoria Ator-Rede reforça a ideia de valoração, pois permite que os elementos não humanos, como os dados de pesquisa, tenham o mesmo valor que os elementos humanos dentro da rede. Ou seja, adota-se a noção de simetria generalizada, cujo pressuposto atribui agência para elementos não humanos (tradicionalmente atribuída apenas aos humanos) em uma dada rede (LATOUR, 1992). A agência dos dados está na possibilidade de serem encontrados e possuírem as condições de serem reutilizados, alterando efetivamente a dinâmica da pesquisa científica.

O campo de pesquisa se deu no âmbito das Ciências Sociais, conforme já citado e foi operacionalizado em três instâncias. Na primeira instância foram analisados os *frameworks* (quadros conceituais) identificados na literatura que apontam as competências essenciais para que pesquisadores possam gerenciar dados de pesquisa. A partir da análise desses quadros conceituais, foram extraídas as habilidades comuns aplicáveis ao letramento para o reuso de dados.

Na segunda instância adotou-se os conjuntos de dados (*datasets*) abertos já disponíveis sobre a experiência empírica de reuso de dados, dividida em três estudos: (i) um com a participação de 361 cientistas de instituições americanas que relata o comportamento dos pesquisadores em compartilhar dados (KIM, 2016 [*datasets*]); (ii) outro com 1.528 respostas de pesquisadores sobre seu comportamento em reusar dados (KIM, 2017 [*datasets*]) – os resultados dessas investigações estão disponíveis em repositórios de dados tais como o Inter-university Consortium for Political and Social Research (ICPSR)⁴; o (iii) terceiro conjunto de dados que foi utilizado é resultado de um estudo com 1.162 pesquisadores que examinou as atitudes e os comportamentos destes no que diz respeito ao compartilhamento dos seus dados e a reutilização de dados abertos em sua própria pesquisa (BERGHMANS *et al.*, 2017 [*datasets*]; LEIDEN UNIVERSITY; ELSEVIER, 2017). Essas análises visaram identificar se há entre as percepções dos pesquisadores um apontamento sobre dificuldades e/ou falta de habilidades para reusar os dados de pesquisa disponíveis. Os conjuntos de dados (*datasets*) podem ser considerados dados secundários na acepção de Mattar (1996, p. 48) que considera tais dados como sendo “aqueles que já foram coletados, tabulados, ordenados e, às vezes, até analisados e que

⁴ O ICPSR é um consórcio internacional com mais de 750 instituições afiliadas no mundo que está na liderança em gestão de dados em Ciências Sociais, disponibilizando mais de 250 mil arquivos de dados de pesquisas (INTER-UNIVERSITY CONSORTIUM FOR POLITICAL AND SOCIAL RESEARCH, 2017).

estão catalogados à disposição dos interessados”.

A terceira instância de análise foi a dos ambientes virtuais de aprendizagem (AVAs), que se propõem a ofertar letramento em gestão de dados. Para esse conjunto de recursos formado por objetos digitais de aprendizagem (vídeos, *e-books*, listas de exercícios) foram levantadas quais são as competências para o letramento em reuso de dados que estão sendo abordadas e se essas competências têm relação com as características identificadas nas análises dos *frameworks*. Todas essas etapas estão detalhadas no Capítulo 5 – Metodologia da Pesquisa.

No que se refere aos recursos que foram analisados nos AVAs, faz-se importante esclarecer que, ainda que a tecnologia, constituída das mídias empregadas, seja um ator relevante que viabiliza o processo de ensino-aprendizagem à distância, seus elementos não foram desdobrados nesse trabalho. Da mesma forma, o presente estudo não envolveu questões ligadas à análise das interfaces do AVAs, como por exemplo, navegabilidade, adaptabilidade e usabilidade (PEREIRA; SCHMITT; DIAS, 2007). A contribuição desta pesquisa na perspectiva dos AVAs se dá na qualidade de formulação de conteúdo, pois tem como prerrogativa de função a “atuação como um especialista no domínio do conhecimento responsável pela preparação do programa didático, criação e seleção de conteúdo” (PEREIRA; SCHMITT; DIAS, 2007, p. 17). A seguir são apresentadas sob quais condições esse tema se configurou um problema de pesquisa, sendo apresentados os pressupostos e a tese que direcionaram este trabalho.

1.3 PROBLEMÁTICA DE PESQUISA, PRESSUPOSTOS E TESE

Manifestos de diversos países e organismos supranacionais, como a Organização para Cooperação e Desenvolvimento Econômico – OCDE (2007), criaram requisitos de publicidade e acessibilidade dos dados resultantes das pesquisas que financiaram, buscando conferir transparência aos investimentos públicos. O manifesto com os Princípios e as Diretrizes da OCDE (2007) para acesso a dados de pesquisa de financiamento público forneceu recomendações sobre o acesso aos dados criados com financiamento público para os 30 governos membros participantes, incluindo os Estados Unidos.

O United States National Institutes of Health (NIH) criou sua política de acesso em 2003 (NATIONAL INSTITUTES OF HEALTH, 2003) e em 2007 foi aprovada uma lei nacional que exigia da NIH o acesso aberto de todas as pesquisas financiadas pelo instituto, sendo que todos os artigos deveriam ser depositados no repositório institucional (RI) PubMedCentral (GRIGG, 2016).

Em 2012, uma petição pública com 65.704 assinaturas exigia o acesso aberto para todas as agências federais norte-americanas que financiavam pesquisas científicas (WE THE PEOPLE, 2012). A resposta da Casa Branca⁵ foi um requerimento para todas as agências federais que têm mais de US\$ 100 milhões em despesas de pesquisa e desenvolvimento solicitando a elaboração de políticas para disponibilização de resultados das pesquisas financiadas pelo governo federal gratuitamente, no prazo de 12 meses após a publicação original, inclusive dos dados de pesquisa (OFFICE OF SCIENCE AND TECHNOLOGY POLICY, 2013). Desde então, 16 departamentos e agências federais responsáveis por mais de 98% das despesas federais de pesquisa completaram seus planos de acesso público e começaram a tomar medidas para fornecer acesso público gratuito a publicações acadêmicas e dados digitais resultantes de pesquisa financiada (CENDI, 2016). Diversos financiadores privados também adotaram políticas em torno de dados de pesquisa (BILL & MELINDA GATES FOUNDATION, 2015; OPEN RESEARCH FUNDERS GROUP, 2016).

A principal instituição norte-americana de fomento à pesquisa, denominada National Science Foundation (NSF), e também a Europa por meio do programa Horizon 2020 (H2020) envolvendo todos os países europeus, já exige planos de gestão de dados como contrapartida da disponibilização de recursos financeiros (EUROPEAN COMMISSION, 2016a; NATIONAL SCIENCE FOUNDATION, 2007).

Destarte, as leis e as políticas em níveis nacionais e internacionais têm influenciado fortemente os dados de pesquisa (BRINEY; GOBEN; ZILINSKI, 2015). Entretanto, ainda que haja cada vez mais a exigência pela abertura de dados, pesquisadores enfrentam diversas dificuldades e desafios diante da prática de disponibilidade e de reuso (AVENTURIER; ALENCAR, 2016).

No Brasil, a título de exemplo, segundo a Rede Nacional de Ensino e Pesquisa (2017b), “as iniciativas de promoção do acesso aberto a dados de pesquisa são incipientes, desarticuladas e pouco padronizadas, restringindo-se a poucas ações encabeçadas por algumas instituições ou grupos de pesquisa” (REDE NACIONAL DE ENSINO E PESQUISA, 2017b, *online*). Ademais, a comunidade internacional Research Data Alliance (RDA)⁶ registrou, no ano de 2017, uma baixa participação brasileira, considerando que apenas 2% dos membros da

⁵ A Casa Branca (tradução literal do inglês: *White House*) é a residência oficial e principal local de trabalho do Presidente dos Estados Unidos, sendo, ao mesmo tempo, a sede oficial do poder executivo naquele país. (https://pt.wikipedia.org/wiki/Casa_Branca).

⁶ A RDA foi criada em 2013 com financiamento da Comissão Europeia, do NSF (Estados Unidos) e do departamento de inovação do governo australiano, com o objetivo de construir a infraestrutura social e técnica para permitir o compartilhamento aberto de dados científicos nas mais diversas áreas do conhecimento. Conta com mais de 5900 membros de 129 países (dados de agosto/17). <https://www.rd-alliance.org/node/51727>.

organização estão localizados na América Latina (RESEARCH DATA ALLIANCE, 2017). Outro estudo de Veiga (2017) mostrou que pesquisadores brasileiros também desconhecem a existência dos recursos de dados abertos. Este desconhecimento reforça a falta de letramento informacional em dados do pesquisador na gestão dos dados de pesquisa. Não conhecer e não saber como utilizar esses recursos impede a possibilidade de intercâmbio de informações entre grupos de pesquisa em determinadas áreas do conhecimento, diminui a visibilidade dos resultados das pesquisas em nível institucional, impede a democratização do acesso aos dados, reduzindo a eficiência do investimento e a transparência na prestação de contas do fomento público à atividade científica (REDE NACIONAL DE ENSINO E PESQUISA, 2017a).

Uma busca preliminar de Rocha (2017), no diretório Registry of Research Data Repositories (re3data.org⁷), um indexador que mantém o registro global de repositórios de dados de pesquisa, também identificou o registro de apenas seis repositórios brasileiros, sendo a metade deles resultantes de cooperação internacional com países como Estados Unidos e Inglaterra.

Impulsionada por esse cenário, a Rede Nacional de Ensino e Pesquisa (RNP) e o Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict) entenderam a importância na concepção de estratégias visando à articulação de atores e de instituições nacionais, estabelecendo assim o Programa Brasileiro de Acesso Aberto aos Dados de Pesquisa – PBAADP (REDE NACIONAL DE ENSINO E PESQUISA, 2017a). Em março de 2017 a RNP e o Ibict abriram uma chamada para financiar um grupo de pesquisa em Acesso Aberto aos Dados de Pesquisa (AADP), cujo objetivo é executar atividades que contribuam para a identificação de práticas, o mapeamento de requisitos e a prototipação de sistemas que facilitem a disseminação de informações científicas.

Por outro lado, apesar da Lei nº 12.527/2011, conhecida como Lei de Acesso à Informação (LAI), que regulamenta o direito constitucional de acesso às informações públicas no Brasil (BRASIL, 2011, *online*) ser a primeira lei de acesso público no mundo que prevê dados abertos, pois exige que os dados de governo sejam publicados em formatos abertos e não proprietários, não há ainda no país outros mandatos específicos que exijam o depósito de dados de pesquisa (SAYÃO; SALES, 2014). Dentre a legislação (Decretos, leis, pareceres, portarias, entre outros) da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Capes,

⁷ A maior parte desses recursos indexados no re3data.org são repositórios para depósito de dados de instituições de pesquisas e de universidades. A pesquisa no re3data.org pode ser feita por um conjunto de critérios, como tipo de acesso, tipo de conteúdo, estrutura de dados, padrão de metadados utilizado, licença de uso de dados, país de origem, idioma, tipo de versionamento, área do conhecimento, *software* utilizados para o repositório, dentre outros.

principal órgão de fomento a pós-graduação no país, foram identificados 70 itens que regulam de alguma forma a concessão de financiamento direto para pesquisadores na pós-graduação (BRASIL, 2017a, *online*). Nenhum desses itens, no entanto, cita a necessidade de elaboração de um plano de gestão e, sobretudo, do compartilhamento de dados como contrapartida ao financiamento de pesquisa.

Apenas a Fundação de Amparo à Pesquisa do Estado de São Paulo – FAPESP (2018) considera o depósito dos dados para alguns tipos de pesquisa, pois:

reconhece a importância da gestão adequada dos dados de pesquisa como parte essencial das boas práticas de pesquisa. Para tanto, essa agência considera necessário que os dados resultantes de projetos financiados pela Fundação sejam gerenciados e compartilhados de forma a garantir o maior benefício possível para o avanço científico e tecnológico. (FAPESP, 2018, *online*).

Sendo assim, para determinadas modalidades e chamadas, o documento Plano de Gestão de Dados faz parte dos anexos obrigatórios de uma proposta de financiamento submetida à FAPESP (2018).

Um dos atores do processo de comunicação científica que tem fomentado a exigência de abertura de dados são os Editores de publicações científicas, por meio de exigência do depósito de dados que embasam os estudos como condição para a aceitação dos trabalhos para publicação, dados esses coletados e utilizados para análise, tratamento e condensação em forma de resultados divulgados nos textos científicos (COPE; KALANTZIZ, 2014). Os Editores exigem que os dados sejam disponibilizados em repositórios próprios ou em repositórios externos (CARVALHO, 2016). Essa exigência tem se intensificado porque com a abertura dos dados de pesquisa é possível o rastreamento, a validação e a transparência da pesquisa, além de viabilizar a sua reprodutibilidade (LEIDEN UNIVERSITY; ELSEVIER, 2017; LEONELLI, 2013; OPEN SCIENCE AND RESEARCH INITIATIVE, 2014).

Apesar da exigibilidade em compartilhar os dados de pesquisa, há várias barreiras de diversificados aspectos como questões socioculturais, tecnológicas, políticas, institucionais, econômicas, de aspectos éticos e legais, como o direito autoral e a proteção intelectual que impedem ou inibem o compartilhamento dos dados pelo pesquisador (BUENO DE LA FUENTE, 2016; FANIEL *et al.*, 2013; FANIEL; MAUTHNER, 2012; ZIMMERMAN, 2011). Ainda que essas barreiras sejam transpostas por meio de políticas de incentivo ou mesmo por força dos mandatos, os dados compartilhados apenas cumprirão seu propósito se de alguma forma forem reutilizados: seja por comparações, verificações ou novas análises em diferentes contextos ou domínios.

O compartilhamento dos dados e a prática de reuso por parte dos pesquisadores pressupõe uma série de interações entre elementos humanos e não-humanos, desde resultados de laboratórios até os arquivos e os dispositivos eletrônicos, passando pelas teorias e políticas da construção (social) da Ciência, nas quais os pesquisadores se norteiam. Somadas à essas relações, acrescentam-se as dificuldades em termos de letramento informacional em dados do pesquisador.

Um recente relatório, que apresenta a perspectiva dos pesquisadores, foi publicado pela Editora Elsevier em conjunto com o Centro de Estudos em Ciência e Tecnologia da Universidade de Leiden (LEIDEN UNIVERSITY; ELSEVIER, 2017). Esse relatório apresenta os resultados de uma pesquisa realizada com acadêmicos do mundo todo, que obteve 1.162 respostas de várias disciplinas do conhecimento. Neste estudo, 73% dos entrevistados concordam que ter acesso aos dados de outros pesquisadores pode ser benéfico, no entanto 34% admitem que nunca publicaram seus dados, porém, 64% destes afirmam que os publicariam (LEIDEN UNIVERSITY; ELSEVIER, 2017). **Contudo, o mesmo relatório mostra que muitos pesquisadores afirmaram que se sentem despreparados, sendo que 41% deles reconheceram que não receberam treinamento suficiente para atender os requisitos em gestão de dados.** O relatório destaca a necessidade de mais orientação para acadêmicos que ainda estão confusos sobre dados abertos.

Outro relatório, que sintetiza as respostas de dois mil estudantes de várias partes do mundo sobre dados abertos de pesquisa (DIGITAL SCIENCE; FIGSHARE, 2016), mostra que mais da metade dos acadêmicos que compartilharam seus dados tinham dúvidas quanto às condições de proteção de propriedade intelectual. Este estudo também revela que é necessária mais orientação para garantir maior clareza em lidar com tal situação, já que 60% dos que depositaram os dados não sabiam quais eram essas condições de proteção e mais da metade afirmaram que gostariam de ter mais preparo para estar em conformidade com os demais requisitos das políticas de compartilhamento (DIGITAL SCIENCE; FIGSHARE, 2016).

No estudo de Tenopir *et al.* (2015a) que avalia a percepção dos pesquisadores sobre suas habilidades em gestão de dados como um todo, a falta de habilidades aparece como principal barreira ao compartilhamento de dados, apresentando diferenças significativas inclusive por regiões geográficas (TENOPIR *et al.*, 2015a). Os pesquisadores da África e da América do Sul estão mais preocupados com a falta de habilidades para disponibilizar seus dados. A pergunta “Eu tenho habilidades insuficientes para disponibilizar meus dados” (TENOPIR *et al.*, 2015a, tradução nossa, pergunta 13b.11) foi respondida de forma afirmativa por 23,1% dos africanos e por 23,8% dos sul-americanos entrevistados. Esse percentual indica

esta barreira como significativamente mais alta do que a média das demais barreiras, que foram indicadas por 13,3% das respostas (TENOPIR *et al.*, 2015a).

Um relatório publicado pela Online Computer Library Center (OCLC)⁸ investigou a perspectiva de algumas instituições de ensino sobre a gestão de dados de pesquisa, dentre elas: a Universidade de Edimburgo (Reino Unido), a Universidade de Illinois em Urbana-Champaign (EUA), a Monash University (Austrália) e a Wageningen University & Research (Países Baixos), além de outras universidades na América do Norte, Europa e Austrália. O relatório apresentou as influências que conduziram as escolhas destas instituições no contexto de dados de pesquisa e o que elas enfrentaram na construção ou aquisição de capacidade em gestão de dados. Dentre as três necessidades de apoio institucional que essas universidades perceberam como necessárias em relação à gestão de dados, duas delas estão relacionadas com a ajuda em melhorar as habilidades dos pesquisadores (BRYANT; LAVOIE; MALPAS, 2017).

De tal cenário, a problemática deste estudo foca, portanto, na falta de Letramento Informacional dos pesquisadores na gestão dos dados de pesquisa. Tenopir *et al.* (2011) demonstraram a partir de um estudo com 1.300 cientistas que a principal barreira para o compartilhamento e reuso dos dados era a falta de letramento em dados dos pesquisadores. Outra pesquisa de 2012 mostrou que nenhum dos pesquisadores estudados tinham recebido uma capacitação formal em gestão de dados e eles não estavam satisfeitos com o seu nível de *expertise* (JAHNKE; ASHER; KERALIS, 2012). Para Kim (2013, p. 501, tradução nossa), os pesquisadores raramente tinham “as habilidades ou os recursos para preparar os seus dados para o compartilhamento público”.

A partir da problemática entre as exigências de compartilhamento e a falta de capacidade do pesquisador na gestão dos dados de pesquisa (KOLTAY, 2017a; CARLSON *et al.*, 2011; MAYBEE *et al.*, 2015, JEFFRYES; JOHNSTON, 2013) por um lado, e por outro lado o papel dos dados que se configuram como um ator relevante no contexto da produção científica, fortemente apoiado em tecnologias da informação; as potencialidades inexploradas pela falta de reuso de dados em uma cultura de ciência aberta *versus* a lacuna de competência informacional para apropriação e de reuso dos dados já disponíveis, esta pesquisa visa responder a seguinte questão:

⁸ A OCLC é uma cooperativa global de bibliotecas, com países participantes de todos os continentes, como Alemanha, Angola, Brasil, Canadá, Chile, Estados Unidos, Itália, Reino Unido dentre outros, sem fins lucrativos, que oferece serviços de tecnologia compartilhada, pesquisas originais e programas comunitários para que as bibliotecas possam contribuir melhor com o ensino, a pesquisa e a inovação.

Quais são os requisitos necessários em ambientes virtuais de aprendizagem para que os pesquisadores desenvolvam competências mínimas de letramento informacional para reuso de dados de pesquisa nas Ciências Sociais, na abordagem da Teoria Ator-Rede?

Ambientes virtuais de pesquisa com acesso aberto aos dados científicos estão disponíveis nos repositórios internacionais e há um tímido movimento nacional em prol do acesso aos dados de pesquisa no Brasil. Entretanto, pressupõe-se que os pesquisadores não estejam habituados e capacitados para a apropriação e reuso desses dados. Promover o letramento em dados e dos ambientes virtuais que os armazenam, a partir de critérios aderentes às necessidades dos pesquisadores, permitiria potencializar o compartilhamento e o reuso dos dados abertos de pesquisa.

Parte-se então dos seguintes pressupostos de pesquisa:

- **Pressuposto 1:** Considerando a sobrecarga de informações do mundo digital de hoje, que, segundo Creaser (2011), também se verifica na produção científica, determinar a finalidade, a qualidade e confiabilidade dos recursos disponíveis é essencial para encontrar informações relevantes aos objetivos dos pesquisadores. Para tanto, é essencial que os pesquisadores tenham o letramento informacional em dados que lhes permitirá determinar, efetivamente, a credibilidade na localização, no acesso e no reuso de recursos digitais. Com a crescente exigência de depósito de dados pelas revistas e a tendência de que um serviço em acesso aberto aos dados de pesquisa seja disponível no Brasil nos próximos anos, presume-se que é preciso que os pesquisadores sejam preparados para essa nova realidade.

O reuso de dados disponíveis é prejudicado pela falta de competências em letramento informacional em dados pelos pesquisadores em identificar, localizar, acessar e reutilizar os dados de pesquisa, especialmente nas Ciências Sociais, deixando assim de usufruir de todo potencial de geração de conhecimento contido nos dados disponibilizados. Viabilizar um modelo para o letramento em dados no ambiente virtual pode alterar essa dinâmica.

- **Pressuposto 2:** A biblioteca acadêmica é o lócus adequado para promover o letramento da gestão, do compartilhamento e do reuso de dados de pesquisa, considerando seu papel social e de responsabilidade com a geração de conhecimento e não apenas com o depósito do mesmo. A biblioteca acadêmica tem, neste contexto, a oportunidade de ser mais

proativa e atuar como um recurso de apoio efetivo nas atividades-fim da universidade pública: o ensino, a pesquisa e a extensão.

- **Pressuposto 3:** A nova descrição do papel dos dados, como forma de simplificação da complexidade, revela a sua justaposição em relação aos outros elementos, demonstrando como a sua atuação modifica a estrutura da rede. Pressupõe-se, portanto, que os dados de pesquisa que são produzidos e esquecidos, desorganizados e não acessíveis nos arquivos pessoais dos pesquisadores ou das instituições, na medida em que são tratados, padronizados, preservados e disponíveis de forma gratuita, constituem um novo actante para se fazer Ciência, a partir do seu reuso.

Para Latour (2000), as inscrições são os alicerces dos produtos científicos, que são os recursos que estão fora dos textos, ou os “recursos que provêm de instrumentos, animais, figuras [...], do lugar de onde saíram: a natureza ou do laboratório” (LATOURE, 2000, p.101). Neste contexto, os dados de pesquisa, se configuram como as inscrições, pois fazem parte da sustentação dos resultados da própria pesquisa. Desta forma, sob a as lentes da Teoria Ator-Rede (LATOURE, 2001), pode-se pressupor que os próprios dados de pesquisa constituem um actante para a investigação e a geração de conhecimento científico. Sendo assim os dados de pesquisa deixam de figurar como subprodutos das atividades de pesquisa e se tornam um foco de grande interesse para todo o mundo científico (SAYÃO; SALES, 2016).

Portanto, acredita-se que a falta de letramento em dados dos pesquisadores diminui o valor potencial de reuso desses dados. Os pressupostos defendem que há requisitos e competências que propiciam o reuso dos dados. De outra parte a Biblioteca Acadêmica é um ator relevante que intermedia esse processo de aprendizagem, por estar diretamente relacionada a atividade fim dos(as) pesquisadores. Neste cenário, **defende-se a tese de que ambientes virtuais de aprendizagem podem ser recursos de suporte à Pesquisa e ampliam a possibilidade de reuso de dados se ofertarem o letramento informacional aos pesquisadores a partir de um conjunto de requisitos aplicáveis ao reuso, principalmente aos critérios que atestem a proveniência, a qualidade e a confiabilidade dos dados secundários. Logo, determinar tais requisitos torna-se condição para o desenvolvimento de competências mínimas para o letramento Informacional.**

O ambiente deste letramento é influenciado por constructos como a rede da Comunicação Científica, os Aspectos Regulatórios, a prática da Ciência Aberta, o desenvolvimento de Recursos Tecnológicos e o paradigma moderno da *eScience*, cujas relações estão estruturadas como proposta maior desta pesquisa – ver Figura 10, Capítulo 4.

Sequencialmente estabelecem-se os objetivos da pesquisa.

1.4 OBJETIVOS DA PESQUISA

Diante dos desafios anteriormente registrados nesta seção, apresentam-se os objetivos geral e específicos que contemplam esta Tese.

1.4.1 Objetivo Geral

Propor, a partir de elementos da Teoria Ator-Rede, requisitos para um ambiente virtual em letramento informacional para reuso de dados nas Ciências Sociais, identificando as competências mínimas para este reuso.

1.4.2 Objetivos Específicos

Para viabilizar o objetivo geral proposto, os seguintes objetivos específicos (OE) são apresentados:

- a) OE1 - Caracterizar ambientes virtuais de pesquisa formados pela infraestrutura de coleta, de armazenamento, de serviços, de disponibilidade e de reuso de dados de pesquisa com base na literatura;
- b) OE2 - Levantar, nos quadros conceituais (*frameworks*) de letramento informacional em dados, os requisitos propostos para o reuso de dados;
- c) OE3 - Identificar as perspectivas e as dificuldades dos pesquisadores no reuso de dados de pesquisa, por meio de dados secundários (*datasets*);
- d) OE4 - Identificar as competências essenciais para o reuso de dados no letramento proposto pelos recursos de aprendizado *online*, disponíveis nos ambientes virtuais de aprendizagem;
- e) OE5 - Comparar as características dos quadros conceituais com as perspectivas dos pesquisadores e com as competências nos recursos de aprendizagem *online*, propondo um conjunto de requisitos com vistas ao reuso de dados nas Ciências Sociais;
- f) OE6 - Validar os requisitos propostos por meio da verificação de aderência às melhores práticas da área recomendadas pelo World Wide Web Consortium (W3C).

A seguir apresentam-se as justificativas da pesquisa.

1.5 JUSTIFICATIVAS TEÓRICO-PRÁTICAS E MOTIVAÇÃO PESSOAL

Alguns princípios atrelados aos dados abertos e que podem ser usados como elementos de justificativa do desenvolvimento desta pesquisa são: a diversidade de opiniões a partir de novos tipos de pesquisas, a verificação de resultados prévios, a promoção de pesquisa interdisciplinar, interinstitucional e internacional (UHLIR; SCHRÖDER, 2007), a indução da colaboração e da transparência dos dados (ORGANIZATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT, 2015), a possibilidade do reúso de dados gerados (CURTY, 2016a). Ainda, a atuação em rede, por meio do compartilhamento dos dados, permite aos pesquisadores alcançarem resultados que sozinhos não seria possível. As possibilidades se ampliam diante do compartilhamento de recursos, tanto em termos financeiros, como do tempo e do espaço.

Como exemplo da colaboração em rede destaca-se o maior programa de pesquisa científica já criado na Europa, o Horizon 2020 (H2020), cujo investimento para o período 2014 e 2020 foi de 80 bilhões de Euros (EUROPEAN COMMISSION, 2014), que tem como exigência a garantia de publicidade em acesso aberto dos resultados das pesquisas financiadas. Esta exigência é uma das principais âncoras do programa, sendo prevista no regulamento com um critério mandatório para a participação e para o pagamento de bolsas aos candidatos (EUROPEAN COMMISSION, 2016a, p. 2016-2019; 2016b).

Outras entidades a serem citadas são a American National Science Foundation (NSF) e o United States National Institutes of Health (NIH), o National Institute of Standards and Technology (Norte-americanos), a Comissão Europeia, o Australian Department of Innovation, que juntos formam o Research Data Alliance (RDA), um esforço internacional que visa construir a infraestrutura social e técnica para permitir o compartilhamento aberto de dados. A aliança reuniu, em novembro de 2017, 6.300 membros de 132 países (RESEARCH DATA ALLIANCE, 2017).

Esse, portanto, é um tema de destaque entre as instituições de financiamento e pesquisa em várias localidades do mundo: em dezembro de 2016, outras oito importantes agências de financiamento, dentre as quais a Bill & Melinda Gates Foundation, anunciaram o lançamento do Open Research Funders Group (ORFG), uma parceria concebida para aumentar o acesso aberto aos resultados de pesquisa (OPEN RESEARCH FUNDERS GROUP, 2016). Com um investimento de quase 5 bilhões de dólares anuais, o ORFG visa aumentar o compartilhamento aberto de artigos e dos dados de pesquisa, acelerando o ritmo e reduzindo as lacunas na partilha de informação, incentivando a inovação e promovendo a reprodutibilidade de dados (OPEN

RESEARCH FUNDERS GROUP, 2016).

Diante de tais iniciativas, depreende-se a notória relevância dos investimentos na promoção da Ciência Aberta, não apenas para dar visibilidade aos resultados das pesquisas, mas também na disponibilidade dos dados que fundamentam estes resultados em artigos e textos científicos financiados por diversas instituições. A Ciência Aberta, portanto, representa uma nova abordagem para o processo científico com base no trabalho cooperativo e em novas formas de difusão de conhecimentos por meio de tecnologias digitais e das novas ferramentas colaborativas (EUROPEAN COMMISSION, 2016c).

A motivação para esse tipo de pesquisa, do ponto de vista das políticas com força de lei, está na necessidade de adequação do pesquisador ao novo modelo de pesquisa científica que vem sendo discutido no mundo acadêmico, buscando conformidade com atuais requisitos mandatórios de revistas científicas e de políticas institucionais, portanto, quem pretende atuar nesse campo precisa estar habilitado para atender a essas exigências. Do ponto de vista sociotécnico, o acesso aberto aos dados faz parte da Ciência Aberta e traz uma cultura de colaboração que favorece a aceleração do processo de conhecimento científico, dentre outros fatores que beneficiam a rede de comunicação científica. Desmembra-se, a partir deste ponto, a Justificativa em teórica e prática, relevando-se pontos que se julgam essenciais de discussão.

1.5.1 Justificativa Teórica

A produção científica, e a comunicação dos seus resultados, é um processo complexo, heterogêneo e interdependente segundo Borgman (2000; 2007). Entender a dinâmica dessas relações e os desdobramentos da atuação dos atores é essencial para que se compreendam os rumos que a produção e a comunicação da ciência têm tomado (SAYÃO; SALES, 2015b). A Teoria Ator-Rede (TAR) pode ser uma lente para observações necessárias à compreensão dessa dinâmica, pois proporciona a ideia de uma teoria geral das relações entre elementos heterogêneos.

Uma nova descrição do papel dos dados, como forma de simplificação da complexidade, revela a sua justaposição em relação aos outros elementos, demonstrando como a sua atuação pode modificar a estrutura da rede. O reuso desses dados, quando aplicado os conceitos da TAR, possibilita a abertura de uma caixa-preta. “A caixa preta contém aquilo que já não precisa ser repensado, coisas cujo conteúdo tornou-se uma questão de indiferença” (LATOUR; CALLON, 1981, p. 284). Latour (2000) descreve as implicações e as circunstâncias da construção de fatos

científicos e de artefatos ao ponto de se tornarem caixas pretas, ou seja, o momento em que esses recursos científicos estão acabados, em que toda a sua complexidade está encapsulada em um conceito ou em um artefato de forma que se torna inquestionavelmente aceito. As caixas-pretas são o último estágio da construção científica. Quaisquer questionamentos ao objeto estabilizado custam muito, discordar e argumentar, abrir a caixa-preta é, para Latour (2000, p. 116), “uma tarefa cara, quiçá inexecutável”. A importância da disponibilidade dos dados de pesquisa se dá, então, principalmente pela possibilidade de compreender o processo de geração daquele conhecimento, sem obscuridades.

Por outro lado, a disponibilidade de dados para reuso vem sendo estimulada em diversas áreas por causa da preocupação crescente com a reprodutibilidade de pesquisas, a fim de que seja possível verificar a precisão e a relevância dos resultados. Recentemente a Real Academia de Artes e Ciências da Holanda recomendou em seu relatório o aumento do rigor metodológico e o apoio aos pesquisadores interessados em reproduzir trabalhos para verificação de resultados, além de reforçar o treinamento em tópicos relacionados ao manejo de dados e à análise estatística (ROYAL NETHERLANDS ACADEMY OF ARTS AND SCIENCES, 2018). Isso porque, segundo este relatório, havia um problema de reprodutibilidade que exige cada vez mais a disponibilidade dos dados comprobatórios dos resultados. Como exemplos foram listadas as tentativas de confirmar os resultados de 53 estudos pré-clínicos com potencial reuso para novos medicamentos contra o câncer, sendo que apenas 11% dos resultados foram confirmados. Outra tentativa de validar dados em 67 projetos de pesquisa sobre medicamentos só teve sucesso em 25% dos casos. Igualmente, em uma área da psicologia experimental só se conseguiu confirmar 36 resultados obtidos dentre 100 artigos avaliados (ROYAL NETHERLANDS ACADEMY OF ARTS AND SCIENCES, 2018).

Para Sayão e Sales (2015a), a disponibilidade e o reuso de dados se complementam. A falta da disponibilidade dos dados inviabiliza a preservação digital dos mesmos, que ficam perdidos nos discos rígidos e nos computadores pessoais dos pesquisadores, sendo que se estivessem depositados em ambientes adequados seriam configurados para serem acessíveis no futuro, apesar da obsolescência tecnológica dos formatos.

Já a falta do reuso de dados disponíveis diminui a velocidade da produção científica em algumas áreas do conhecimento (SAYÃO; SALES, 2014). Um estudo realizado com uma amostra de respostas de 595 pesquisadores de diferentes disciplinas e países, evidenciou que a prática de reuso de dados, ou seja, novos estudos que tem como base os dados gerados em pesquisas anteriores, ainda que esteja crescendo, dissemina-se de forma mais recorrente nas Ciências Exatas e Biológicas, enfrentando muita resistência nas Ciências Sociais (CURTY *et*

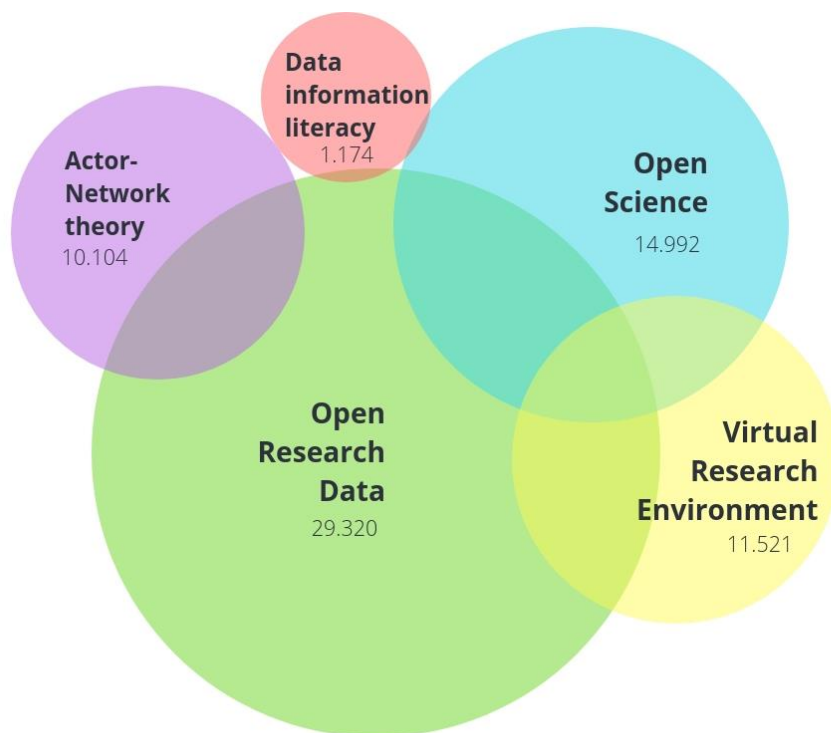
al., 2017).

Os dados do estudo de Curty *et al.* (2017) foram reutilizados a partir de uma pesquisa original com mais de mil questionários respondidos por cientistas de várias partes do mundo (TENOPIR *et al.*, 2015a). Este estudo original foi aplicado pelo grupo de pesquisa do Data Observation Network for Earth (DataONE), cujo projeto é financiado pela National Science Foundation (NSF) (TENOPIR *et al.*, 2015a). Curty *et al.* (2017) destacam as barreiras e os comportamentos inibidores ao reúso de dados das Ciências Sociais, ligados a questões éticas, de má interpretação e de uso incorreto de dados, características que estão relacionadas ao letramento para habilidades e competências em manejo de dados secundários. O Letramento em dados e o letramento informacional são e continuarão sendo tão importantes (CALZADA PRADO; MARZAL, 2013) principalmente devido ao aumento da complexidade do ambiente informacional (NEW MEDIA CONSORTIUM, 2013), sendo essenciais dentre as competências de pós-graduação (DELANEY; BATES, 2015).

De outra parte, para levantar a relevância teórica da pesquisa e sua originalidade, sob o aspecto da produção científica, foi realizado um levantamento bibliográfico preliminar, a partir de cinco grandes conceitos correlatos ao tema: (i) Ciência Aberta, (ii) Dados abertos de Pesquisa, (iii) Letramento informacional em dados, (iv) Teoria Ator-rede e (v) Ambiente Virtual de Pesquisa. A pesquisa foi realizada, com os termos versados para a língua inglesa, nas seguintes fontes de informação: (i) O Portal de Periódicos da Capes, (ii) o Portal EBSCOHost e a (iii) Base de dados referencial Web of Science.

A quantidade de artigos recuperada na busca isolada dos termos é apresentada na Figura 2, que utiliza a proporção gráfica para representar a diferença entre as quantidades recuperadas de cada descritor. O tamanho de cada círculo representa o total de resultados na busca de cada assunto e a diferença de tamanho entre eles a proporcionalidade da recuperação.

Figura 2 - Frequência de conceitos-chaves – busca simples



Fonte: Autoria própria (2019).

Nota-se que o conceito-chave principal desta pesquisa expresso pelos descritores *Open Research Data* (Dados abertos de Pesquisa), com 29.320 resultados, é amplamente discutido na literatura, mesmo quando pesquisado sem combinação com outros conceitos. O descritor *Open Science* (Ciência Aberta), segundo maior resultado, recuperou 14.992 registros, com ampla recuperação portanto, seguido do terceiro descritor *Virtual research environment* (Ambiente Virtual de Pesquisa), com 11.521 resultados, igualmente com recuperação significativa, bem como a dos resultados da busca por *Actor-Network Theory* (Teoria Ator-rede), que totalizaram 10.104 ocorrências. O descritor *Data information literacy* (Letramento informacional em dados) recuperou, no entanto, 1.174 resultados, significativamente menor que os demais dados levantados.

A estratégia de busca foi, na sequência, refinada para uma busca combinada. A base foi constituída dos operadores booleanos (AND) e (OR) e a aplicação de recursos de exatidão na recuperação da informação, como o uso das aspas na combinação dos conceitos entre si. Para essa busca foram consideradas as seguintes combinações:

- Combinação (A): “Open Science” (AND) “Open Research Data”
- Combinação (B): “Open Science” (AND) “Open Research Data” (AND) “Data information literacy”
- Combinação (C): “Open Research Data” (AND) “Data information literacy”
- Combinação (D): “Open Research Data” (AND) “Data information literacy” (AND) “Actor-Network theory”
- Combinação (E): “Virtual Research Environment” (AND) “Open Research Data”

Quando pesquisados de forma combinada entre si e com os outros conceitos norteadores da pesquisa, a quantidade recuperada é baixa, como mostra a Tabela 1.

Tabela 1 - Frequência de Conceitos-Chaves – Busca Combinada

Busca Combinada	Referências Capes	Referências EBSCOHost	Referências Web of Science	Total Referências
Combinação A	163	143	202	508
Combinação B	2	1	1	4
Combinação C	19	116	38	173
Combinação D	6	0	0	6
Combinação E	4	24	36	64
Total referências	194	284	277	755
% de referências por bases	25,70%	37,62%	36,69%	100,00%

Fonte: Autoria própria (2019).

Destaca-se que a combinação B é importante porque demonstra o relacionamento entre a Ciência Aberta e os Dados de Pesquisa em conjunto com o Letramento Informacional em Dados, objeto desta pesquisa. O fato de que haja apenas quatro trabalhos nesta combinação de assuntos (no ano desta pesquisa e nas bases pesquisadas) aponta para a originalidade deste estudo, considerando-se que os resultados das buscas individualizadas atestam revocação substancial dos termos.

A combinação D demonstra que o conceito de Dados Abertos em conjunto com Letramento Informacional em Dados, se combinados com a Teoria Ator-rede, também tem um resultado mínimo, 6 trabalhos. Essa evidência denota a relevância da presente pesquisa, já que aborda tanto o conceito de Letramento Informacional em Dados quanto o conceito da Teoria

Ator-Rede entrelaçados aos temas de Ciência Aberta e de Dados de Pesquisa.

Complementando, foi feito o levantamento de teses e de dissertações correlatas, que foram analisadas pelo título e pelo resumo a fim de se verificar a aderência do trabalho com os termos pesquisados. Foram selecionadas as teses e dissertações mais relevantes, indicadas na Tabela 2, em um total de 16 trabalhos. As referências completas desses trabalhos estão listadas no Apêndice A.

Tabela 2 - Levantamento de Teses e Dissertações

Fontes/ Estratégias de Busca	OasisBR / BDTD	Catálogo de Teses da Capes	Selecionadas
Combinação A			
Dissertações	9	133	4 ⁹
Teses	3	57	3 ¹⁰
Combinação B			
Dissertações	32	0	0
Teses	13	0	0
Combinação C			
Dissertações	31	1	5 ¹¹
Teses	12	0	4 ¹²
Combinação D			
Dissertações	0	0	0
Teses	0	0	0
Combinação E			
Dissertações	8	52	0
Teses	2	30	0
TOTAL			
Dissertações	80	186	9
Teses	30	87	7
TOTAL GERAL	110	274	16

Fonte: Autoria própria (2019).

Nota-se na Tabela 2 que não há qualquer trabalho científico que considere a questão de Letramento Informacional em Dados Abertos de Pesquisa no contexto da Teoria Ator-rede

⁹ APPEL, 2014; CARVALHO, 2018; FIGUEIREDO, 2018; JORGE, 2018; MACHADO, 2015; MAIA, 2018; MONTEIRO, 2017; SOUZA, 2016; SANTOS, 2018;

¹⁰ MARQUES, 2014; MEDEIROS, 2015; PRADO, 2017.

¹¹ OLIVEIRA, 2016; COELHO, 2017; AMORIM, 2013; NOVELLI, 2012; TYGEL, 2016.

¹² SILVA, 2017; COSTA, 2017; ALVES, 2016; SOUSA, 2014.

(combinação D), garantindo, neste momento, a originalidade deste estudo ao usar esses temas de forma combinada, no âmbito da pesquisa *Stricto Sensu*. Dentre todas as teses e dissertações recuperadas, apenas uma abordava o tema Dados de Pesquisa sob a perspectiva da Teoria Ator-Rede, porém, com o foco na propriedade intelectual da autoria científica a partir do compartilhamento de dados de pesquisa (MEDEIROS, 2015). Assim, ainda que haja mais trabalhos científicos abordando o tema Dados de Pesquisa e/ou Teoria Ator-rede, os aspectos tratados não contemplam o Letramento para reuso dos dados, o que reforça o teor de originalidade dessa pesquisa.

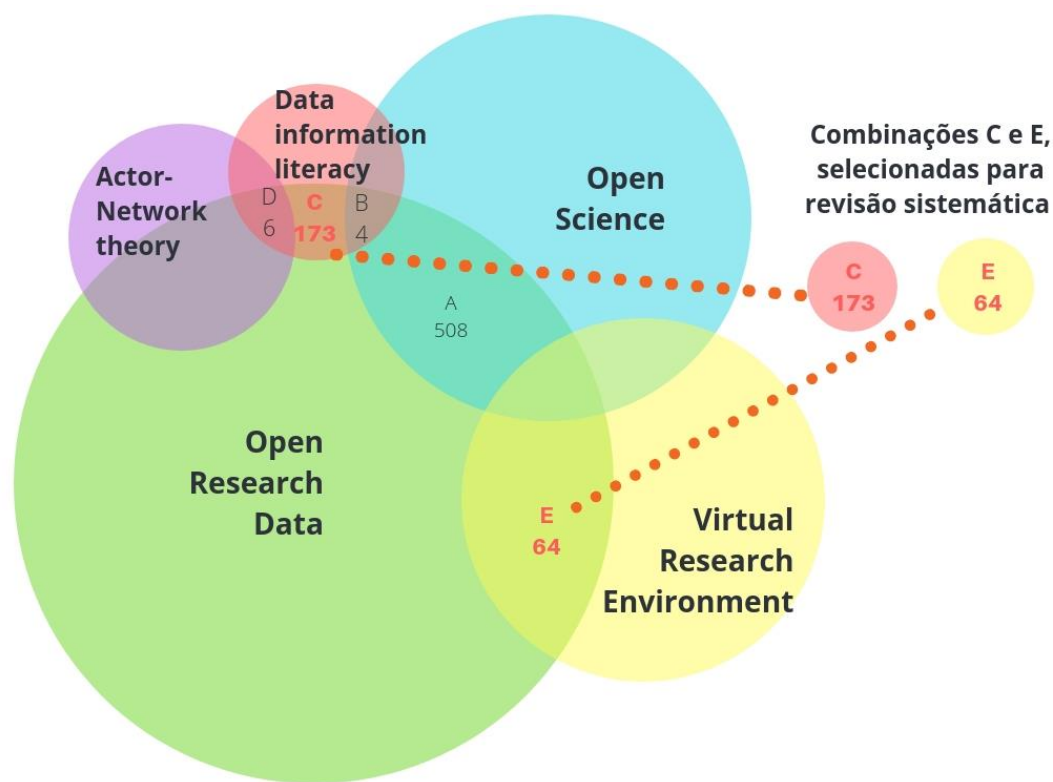
Uma representação gráfica da combinação entre os descritores é apresentada na Figura 3, com a quantidade recuperada e as combinações selecionadas para a revisão sistemática, que foram as Combinação C e E. A quantidade de documentos recuperados em cada busca combinada está presente, igualmente, na Figura 3. Cada termo de busca aparece na proporção da quantidade recuperada quando pesquisados de forma isolada. A intersecção dos termos entre si apresenta o resultado da busca combinada, indicados pelas letras A: 508 resultados; B: 4 resultado; C: 173 resultados; D: 6 resultados e E: 64 resultados. Salienta-se que apenas as combinações C (173 resultados) e a combinação E (64 resultados) foram selecionadas para a revisão sistemática, lembrando-se:

- Combinação (C): “Open Research Data” (AND) “Data information literacy”
- Combinação (E): “Virtual Research Environment” (AND) “Open Research Data”

A combinação A inclui o conceito de “dados de pesquisa”, que também já é considerado na combinação C. Portanto, julgou-se mais relevante os resultados da combinação C, por adicionar aos conceitos “Ciência Aberta” e “dados de pesquisa” o conceito de “letramento informacional em dados”. Por isso optou-se por manter apenas o resultado da combinação C na revisão sistemática, desprezando os resultados da combinação A, já que os resultados das combinações B e D já estavam incorporados, pontificando-se que:

- a Combinação B apresentou apenas 4 resultados, que também estão contidos na Combinação C;
- a Combinação D apresentou 6 resultados, que também está contido na Combinação C;
- a Combinação E apresentou 64 resultados ao adicionar-se o conceito de “ambiente virtual de pesquisa”, por isso essas referências também foram incluídas na revisão sistemática.

Figura 3 - Combinação entre os descritores – quantidade recuperada e selecionada para a revisão sistemática



Fonte: Autoria própria (2019).

Exposto o entendimento da originalidade do tema, a julgar pelos poucos resultados de trabalhos encontrados nas buscas das **Combinações (B)**: “Open Science” (AND) “Open Research Data” (AND) “Data information literacy” e da **Combinação (D)**: “Open Research Data” (AND) “Data information literacy” (AND) “Actor-Network theory”, tanto no levantamento de artigos como nas teses e dissertações, evidencia-se, igualmente, o ineditismo da mesma e a importância teórica do estudo para o Programa de Pós Graduação em Tecnologia e Sociedade (PPGTE), para a Linha de Pesquisa correlata, o Grupo de Pesquisa e o trabalho da Orientação. Sequencialmente trata-se da justificativa prática da pesquisa e a motivação pessoal para a mesma.

1.5.2 Justificativa Prática e Motivação Pessoal

Mesmo que não haja leis no Brasil que determinem o compartilhamento de dados, as revistas científicas já fazem essa exigência, portanto, os pesquisadores precisam estar aptos para

atuar nesses ambientes (CARVALHO, 2016). Após uma investigação preliminar desta pesquisadora em julho de 2017, não foram localizados nos *sites* das bibliotecas acadêmicas de universidades públicas brasileiras, por exemplo, programas de capacitação em competência informacional em dados, como existem para as pesquisas em fontes de informação científicas, como bases de teses, dissertações e artigos. As universidades brasileiras também não têm ainda disponibilizado repositórios para depósito e busca de conjunto de dados de pesquisa. Apenas em dezembro de 2017 o primeiro repositório de uma universidade brasileira foi registrado no re3data.org, a Base de Dados Científicos da Universidade Federal do Paraná (UFPR).

Em pesquisa realizada no *site* re3data.org em janeiro de 2018, que registra e classifica vários repositórios de dados, apenas 7 desses foram listados no resultado da busca com filtro de repositórios de instituições brasileiras: o (i) Brazil Exploration and Production Database, que é o Banco de Dados de Exploração e Produção (BDEP) da Agência Nacional do Petróleo, Gás Natural e Biocombustíveis (ANP); o (ii) WorldClim - Global Climate Data, que pertence a um consórcio cujo Centro de Referência em Informação Ambiental (CRIA) é integrante; o (iii) Global Collaboration Engine – GLOBE, outro consórcio cujo membro brasileiro é o Instituto Nacional de Pesquisas Espaciais (INPE); o (iv) Ibict Dataverse Network, que pertence ao Instituto Brasileiro de Informação em Ciência e Tecnologia em conjunto com o Ministério da Ciência, Tecnologia e Inovação (MCTI); o (v) repositório International Ocean Discovery Program (IODP), que tem a participação da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), (vi) o PPBio Data Repository, repositório de Dados de Levantamentos Biológicos, ligado ao Centro de Estudos Integrados da Biodiversidade Amazônica, e a (vii) Base de Dados Científicos da Universidade Federal do Paraná (UFPR), primeiro repositório de dados de uma universidade registrado no re3data.org.

No Brasil não existe a infraestrutura necessária de armazenagem de dados de pesquisa (ROCHA, 2017), existem dados que estão sendo depositados no exterior em repositórios estrangeiros, disponíveis também na área das Ciências Sociais. Há grande oferta de repositórios de dados disponibilizados em outros países, constituindo fontes de dados para reuso que são listados no diretório re3data.org. Em 2018 estavam registrados 2.032 repositórios em diversas disciplinas. Destes, 598 (29%) cobrem a área de Ciências Sociais e Humanidades. Até o ano de 2017, o Programa Brasileiro de Acesso Aberto aos Dados de Pesquisa, em um esforço conjunto da RNP e do Ibict, ainda buscava identificar as práticas e as iniciativas pelo acesso aberto no âmbito das instituições brasileiras (REDE NACIONAL DE ENSINO E PESQUISA, 2017b). Mesmo com a disponibilidade de dados científicos abertos como mencionado, o principal desafio em promover o reuso desses dados, para Claudia Bauzer Medeiros, do Instituto de

Computação da Universidade Estadual de Campinas (Unicamp) e coordenadora adjunta do programa *eScience* da FAPESP, além de mostrar aos pesquisadores os benefícios e combater os casos de apropriação indevida, é estimular a capacitação da comunidade acadêmica para o preparo e reuso dos dados (ANDRADE, 2018). A Coordenadora ressalta que em alguns países esse tipo de treinamento é um requisito mandatório para a formação de pesquisadores. Na busca pelo letramento em dados, se destaca a necessidade de ferramentas educacionais para que bibliotecários possam usar no ensino de gestão de dados para pesquisadores em diversos campos.

Por isso em países cujo depósito de dados de pesquisa é requisito exigido pelas agências de fomento, como os Estados Unidos, alguns serviços de apoio aos pesquisadores para o manejo dos dados são oferecidos pelas bibliotecas acadêmicas (TENOPIR; BIRCH; ALLARD, 2012; TENOPIR *et al.*, 2015b, TENOPIR *et al.*, 2016). Entretanto, dois estudos compararam a oferta desses serviços entre 2011 e 2014, identificando pouca evolução dos serviços efetivamente disponíveis (TENOPIR; BIRCH; ALLARD, 2012; TENOPIR *et al.*, 2015b). Tal fato se deu principalmente devido à falta de tempo para a capacitação e à falta de apoio institucional da alta direção, contribuindo para a estagnação do nível de serviços ofertados (TENOPIR *et al.*, 2017). Nesse sentido, a capacitação se torna um importante elemento para viabilizar os serviços em gestão de dados de pesquisa direcionados à comunidade acadêmica. Apesar de não haver ainda a exigência formal dos órgãos de fomento nacionais (SAYÃO; SALES, 2014), um estudo sobre a atuação das revistas científicas no contexto de dados abertos mostrou que dentre as 77 revistas brasileiras da área das ciências disponíveis no Directory of Open Access Journals (DOAJ), 15 delas já exigem o depósito de dados, sendo uma delas um periódico exclusivo de dados. Na área de Medicina, das 139 revistas indexadas, 71 fazem a mesma exigência (CARVALHO, 2016).

Para viabilizar a replicação dos procedimentos que levaram às conclusões dos trabalhos a revista *Brazilian Political Science Review*, da Associação Brasileira de Ciência Política, exige a disponibilidade de dados de artigos que utiliza dados quantitativos. Também são requeridos os *codebooks*, ou seja a documentação que identificam as variáveis e as respostas correspondentes (BRAZILIAN POLITICAL SCIENCE REVIEW, [s.d.], *online*).

A publicação no grupo científico Public Library of Science (PLOS) também está condicionada à divulgação dos dados brutos utilizados na pesquisa em repositórios públicos: “Todos os dados e metadados relacionados que fundamentam as descobertas relatadas em um manuscrito submetido devem ser depositados em um repositório público apropriado” (PUBLIC LIBRARY OF SCIENCE, [s.d.], *online*, tradução nossa).

Para lidar com vastas quantidades de dados, a fluência em alfabetização tecnológica ou informática é também necessária (CREASER, 2011). Visando atenuar essas lacunas de letramento, diversas instituições têm criado serviços de apoio aos dados com ações de letramento informacional para viabilizar o acesso aberto e o reuso de dados de pesquisa e, embora não se tenha identificado um exemplo nacional, identificaram-se iniciativas internacionais que já atuam dessa forma (BERMAN, 2013; CARLSON *et al.*, 2011; MACY; COATES, 2016; FRANK; PHARO, 2016; KOLTAY, 2015a; KOLTAY, 2016; KOLTAY, 2017a; KOLTAY, 2017b; MANDINACH; GUMMER, 2013; MANDINACH; GUMMER, 2015).

Por outro lado esta pesquisa se delimita na área das Ciências Sociais, como já explicitado, e estabelece suas relações com as tecnologias, e seus desdobramentos, nos sistemas sociotécnicos e nos processos científicos, objetos de estudo do Programa de Pós-Graduação em Tecnologia e Sociedade – PPGTE, cuja atuação se dá no âmbito interdisciplinar em torno de investigações que abrangem as interações entre Tecnologia e Sociedade. Como pode ser observado na apresentação do Programa, o propósito deste é “pesquisar as transformações que as mudanças tecnológicas provocam nas atividades realizadas pela sociedade no âmbito material e cultural” (UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ, 2018, *online*). Os estudos Sociais da Ciência e da Tecnologia (CTS) há décadas têm constituído um campo do conhecimento que permite a análise dessas relações e dos seus desdobramentos na sociedade (CUTCLIFFE, 2003). O PPGTE investiga, sobretudo, “como essas inovações interferem na vida das pessoas, na sua maneira de trabalhar, aprender, pensar, simbolizar e atuar no mundo” (UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ, [s.d.], *online*).

Logo, o objeto do presente estudo se relaciona com o Programa de Pós Graduação em Tecnologia e Sociedade (PPGTE), por conta das características heterogêneas de sua constituição; das relações dinâmicas em rede, cujos elementos apresentam estreita relação com a Ciência, a Tecnologia e a Sociedade, a saber: os dados abertos de pesquisa – aspecto científico, o letramento informacional – aspecto social, os ambientes virtuais de pesquisa – aspecto tecnológico; se aproxima das dinâmicas interdependentes, próprias dos estudos CTS, visto que a característica multidisciplinar é mais prevalente para todo o campo CTS, embora existam também elementos de inter e transdisciplinaridade na Ciência, Tecnologia e Sociedade (VACCAREZZA, 2011). Um dos pilares dos Estudos de CTS é a crítica sobre o desenvolvimento da Ciência e Tecnologia (C&T), cujo modelo tradicional traz em seu bojo a relação direta de que quanto mais Ciência, mais haveria desenvolvimento tecnológico e assim se geraria mais riquezas, que, por consequência, trariam mais bem-estar social. Dessa forma, a

Ciência teria que prosseguir dissociada da sociedade, autônoma e livre da interferência de valores, com atividades supostamente neutras (LINSINGEN; BAZZO; PEREIRA, 2003). Para Feenberg (2010), os resultados da C&T exercem influência na vida diária das pessoas, e não são neutros por ainda estar sob o controle dos senhores dos sistemas técnicos sendo necessária a participação de outros atores no processo de construção da ciência e, conseqüentemente, dos demais artefatos (FEENBERG, 1995). Feenberg (2010) argumentava que a tecnologia moderna, se administrada em outro contexto social, poderia ser um instrumento de democratização.

Diante de algumas prerrogativas da Ciência Aberta – conceito-chave mais amplo dos temas tratados nesta pesquisa, como o acesso aberto, o compartilhamento e o reúso dos dados – a proposta da Ciência Aberta é uma alternativa de resposta à essas críticas. Albagli (2015) discute, nesse contexto, como a Ciência Aberta está desestabilizando modelos vigentes, ao situar-se frente aos conflitos e contradições relacionados ao conhecimento proprietário e ao conhecimento aberto, entre a socialização e a privatização do conhecimento, da informação e da cultura, demonstrando as relações de saber e poder e postulando maior inserção popular na Ciência (ALBAGLI, 2015).

Outrossim, uma das abordagens da Ciência Aberta é a ampliação da participação cidadã no acesso e nos processos de desenvolvimento científicos (e conseqüentemente tecnológicos) em larga escala, como por exemplo: a participação cidadã na geração do conhecimento científico, a ampliação de canais de disseminação da informação científica, a simplificação da linguagem, a aproximação de um público mais amplo aos resultados de pesquisa (FECHER; FRIESIKE, 2014). Então, a presente pesquisa está aderente aos anseios e questionamentos levantados há décadas no campo dos estudos de CTS, e no PPGTE, sobre a produção científica ainda dissociada da sociedade.

Já na Linha de Pesquisa Tecnologia e Desenvolvimento – TD, a qual esta pesquisadora é ligada, e cujo foco está no desenvolvimento territorial sustentável, o estudo contribui sob o aspecto da Territorialidade – um dos eixos de pesquisa da Linha TD, pois apresenta o contexto do acesso aberto aos dados de pesquisa nos âmbitos do Brasil, de países da América do Norte (Estados Unidos e Canadá) e do bloco europeu. Ao incorporar a questão da Territorialidade, esta não se limita a ideia de espaço geográfico, pois o “território não se reduz então à sua dimensão material ou concreta; ele é, também, um campo de forças, uma teia ou rede de relações sociais” (RAFFESTIN, 1993¹³ *apud* ALBAGLI, 2004, p. 26). Esse campo se projeta no espaço

¹³ RAFFESTIN, Claude. **Por uma geografia do poder**. São Paulo: Ática, 1993.

físico, podendo ser objeto de análise em perspectivas distintas que extrapolam a abordagem geográfica como, por exemplo, as perspectivas: antropológica, cultural, social, econômica, política e jurídica, podendo, portanto, o território ser percebido segundo abordagens específicas (ALBAGLI, 2004).

Destaca-se, por ser um espaço único em sua forma de organização e utilização de recursos, a predominância de que determinadas ideias, valores e crenças conferem à territorialidade a evolução de uma percepção de simples qualidade jurídica para a qualificá-lo como um sistema de comportamento (ALBAGLI, 2004). Nesta perspectiva, o acesso aberto aos dados de pesquisa está apropriado em distintos níveis em diversos países, o que também pode ser observado pelos diferentes entendimentos jurídicos de cada nação em relação ao tema.

Por exemplo, os Estados Unidos já aprovaram há mais de uma década legislações¹⁴ que regulam a abertura de dados de pesquisa financiada com recursos públicos, assim como o Canadá¹⁵ e a Europa¹⁶. No Brasil, a perspectiva política e jurídica para a questão ainda está em nível de projetos, como pode ser visto na seção 2.3.1. Essas distinções revelam a construção social de um determinado aspecto em um dado território (RAFFESTIN, 1993¹⁷ *apud* FLORES, 2006).

A relação mais estreita deste estudo é com o grupo de pesquisa Território – Redes, Políticas, Tecnologia e Desenvolvimento, da UTFPR, que se dedica à compreensão da Territorialidade sob a perspectiva das dinâmicas de diferentes naturezas de redes, da tecnologia e do desenvolvimento que moldam o território.

Em relação à Orientação, esta pesquisa vai ao encontro dos estudos da orientadora, que privilegia o ambiente de geração de conhecimento, a gestão da informação científica e tecnológica, além das práticas colaborativas de compartilhamento do conhecimento. A geração de novos conhecimentos e inferências a partir do reuso de dados de pesquisa se insere nesse contexto.

Como motivação pessoal para a escolha do tema, destaca-se o interesse profissional da pesquisadora que atua em uma universidade pública federal e identificou na literatura a relevância para a comunidade acadêmica à qual serve. Ao realizar as buscas por tendências na

¹⁴ (NATIONAL INSTITUTES OF HEALTH, 2003, 2010; NATIONAL SCIENCE BOARD, 2005; NATIONAL SCIENCE FOUNDATION, 2007, 2010; OFFICE OF SCIENCE AND TECHNOLOGY POLICY, 2013)

¹⁵ (SOCIAL SCIENCES AND HUMANITIES RESEARCH COUNCIL OF CANADA, 1990; INDUSTRY CANADA, 2014; TRI-AGENCY, 2014; RESEARCH DATA CANADA, 2016)

¹⁶ (OECD, 2004, 2007; EUROPEAN COMMISSION, 2006; UK RESEARCH COUNCIL, 2005; MEDICAL RESEARCH COUNCIL, 2006; BIOTECHNOLOGY AND BIOLOGICAL SCIENCE RESEARCH COUNCIL, 2008, THE WELCOME TRUST, 2007, 2017).

¹⁷ RAFFESTIN, Claude. **Por uma geografia do poder**. São Paulo: Ática, 1993.

área de pesquisa instrucional, campo prático de atuação da pesquisadora, a obra recém lançada de Kellam e Thompson (2016), foi a fonte de inspiração e que permitiu perceber a lacuna e a oportunidade de pesquisa. Ao realizar-se o levantamento bibliométrico, reafirmou-se o pioneirismo do tema para o Brasil, condição essencial à uma pesquisa de Doutorado.

1.6 ABORDAGEM METODOLÓGICA DA PESQUISA

A abordagem metodológica da pesquisa fornece, segundo Moreira e Caleffe (2006, p. 42), um “conjunto de lentes” para ver, e entender, o universo pesquisado; determina em realidade o paradigma de pesquisa a ser adotado. O paradigma de pesquisa comporta três elementos; (1) a ontologia – natureza e essência da realidade pesquisada; (2) epistemologia – natureza e tipologia de criação do conhecimento, bem como a relação que se estabelece entre investigadores e o objeto investigado e (3) a metodologia os métodos que permitirão apreender dados e informações para identificar as estratégias, os planos de ação, as técnicas de coleta de dados, para orientar o desenvolvimento de projetos de pesquisa (CRESWELL; CLARK, 2013; MOREIRA; CALEFFE, 2006).

Delimitada na área das Ciências Sociais esta pesquisa adotou desde o seu início uma abordagem ontológica de ordem interpretativista que se caracteriza pelo “descrever e interpretar o fenômeno do mundo em uma tentativa de compartilhar significado com os outros” (MOREIRA; CALEFFE, 2006, p. 61), levando em geral a tratamentos qualitativos dos dados coletados empregando vários métodos e técnicas com “muita leitura e reflexão sobre as obras selecionadas, que tratem de teorias e de conhecimentos já existentes, relativos ao problema de investigação (MARCONI; LAKATOS, 2010, p. 272).

Da questão epistemológica entende-se o conhecimento como intangível, construído socialmente, fundamentado na interação entre diversos atores e situações (MOREIRA; CALEFFE, 2008), justificando o uso de diversos métodos, técnicas e procedimentos para se chegar aos dados necessários ao desenvolvimento da pesquisa.

Quanto ao principal método empregado a pesquisa é bibliográfica, pois extrai seus conceitos da literatura e de documentos publicados (GIL, 2010; MARCONI; LAKATOS, 2017). O objetivo deste tipo de estudo é colocar o pesquisador em contato direto com o que foi escrito sobre um determinado tema (LAKATOS; MARKONI, 2017, p. 200). Os recursos informacionais para atender aos objetivos dessa pesquisa se baseiam em materiais bibliográficos e em dados de pesquisa secundários. Os materiais bibliográficos advieram especialmente da área da Ciência da Informação e das Ciências Sociais. Os dados de pesquisa

secundários foram recuperados em repositório de dados abertos.

Nesse método de pesquisa uma variedade de técnicas, para coleta e tratamento dos dados, pode ser aplicada, incluindo a "análise de conteúdo para extrair generalizações com o propósito de produzir categorias conceituais que possam vir a ser operacionalizada sem um estudo subsequente" (MARCONI; LAKATOS, 2012, p. 71). Desta abordagem este estudo é classificado como predominantemente qualitativo, pois utiliza como técnica a análise de conteúdo tanto das fontes bibliográficas como dos conjuntos de dados a serem reutilizados (GIL, 2010; MARCONI; LAKATOS, 2017), sem contudo descuidar da análise quantitativa, quando necessário.

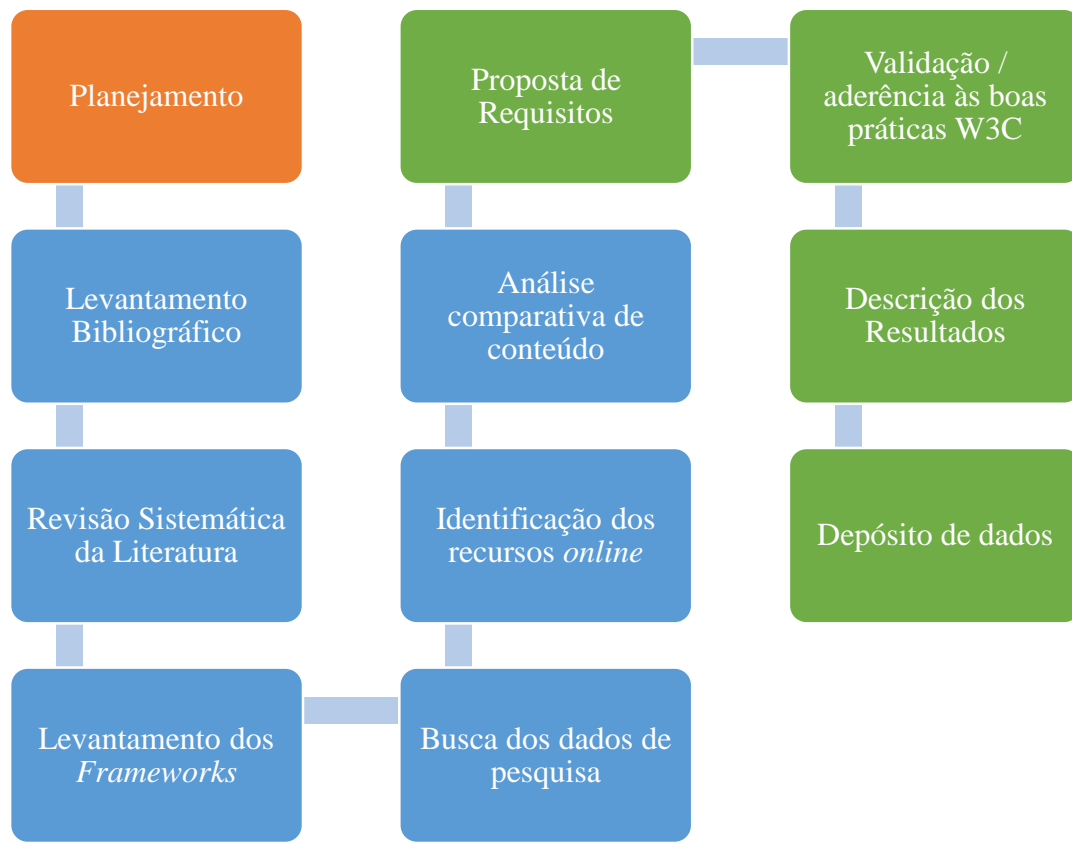
A análise de conteúdo foi fundamentada na abordagem descrita por Bardin (1995), que a conceituou como um conjunto de técnicas de análise de comunicação, viabilizando a partir de “procedimentos sistemáticos e objetivos de descrição do conteúdo das mensagens , a obtenção de indicadores quantitativos ou não, que permitam a inferência de conhecimentos relativos às condições de produção e recepção das mensagens” (BARDIN, 1995, p. 42).

Após o levantamento do *corpus* de pesquisa, a análise de conteúdo foi apoiada pela ferramenta de análises NVivo. As macro etapas desta pesquisa foram realizadas em três fases: (i) o planejamento da pesquisa, o (ii) o desenvolvimento e a (iii) apresentação dos resultados, conforme delinea Gil (2010) e estão apresentadas na Figura 4. Os desdobramentos de cada etapa serão tratados no Capítulo 5.

O planejamento é a etapa anterior a execução da pesquisa propriamente dita, em que se elabora o projeto de pesquisa. Essa etapa se consolidou com a elaboração do projeto da pesquisa e com o plano de gestão de dados. O plano de gestão de dados (Data Management Plan – DMP¹⁸) está disponível no Apêndice D. O desenvolvimento da Pesquisa consistiu nas etapas da levantamento bibliográfico, da análise sistemática, do levantamento dos *frameworks*, da busca dos dados de pesquisa, da identificação dos recursos *online* e da análise comparativa de conteúdo. A etapa final dos resultados passou pela proposta conceitual, a descrição dos resultados e culminará com o depósito dos dados em acesso aberto.

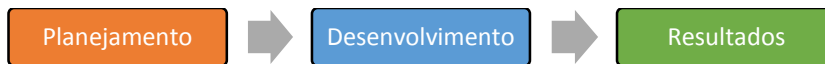
¹⁸ O plano de gestão de dados (Data Management Plan – DMP) é uma atividade em que se prevê o tratamento dos dados em todo o ciclo de pesquisa.

Figura 4 - Planejamento da pesquisa – macro etapas de pesquisa



Fonte: Adaptado de Gil (2010).

Legenda:



Sequencialmente são apresentados os marcos teóricos que embasam as principais teorias abordadas nesta pesquisa.

1.7 EMBASAMENTO TEÓRICO

Os fluxos de informação da comunicação dos resultados de pesquisa abordados por Meadows (1974, 1998), têm sido considerados seminais na literatura sobre a comunicação científica. Sayão e Sales (2012; 2016) têm desenvolvido fortemente a pesquisa no Brasil em como o compartilhamento e o reuso de dados requerem uma *ciber* infraestrutura como ambiente de pesquisa para viabilizar a atividade de gestão e de curadoria de dados, o que confere um novo elemento e altera os modelos de Meadows (1974) introduzindo a necessidade de novas formas de tratamento e gestão dos dados (CARLSON *et al.*, 2011; JAHNKE; ASHER; KERALIS, 2012; TENOPIR *et al.*, 2016).

Koltay (2015a; 2016a; 2017a) evidencia a distinção entre letramento informacional e letramento em dados, detalhando as particularidades e os requisitos inerentes à atividade científica sob a perspectiva dos dados. A gestão e compartilhamento do conhecimento por meio do reuso de dados tem sido a área de concentração dos estudos de Tenopir (2011, 2015), além de abordar como os serviços de apoio ao letramento em dados são essenciais para a comunidade acadêmica. Dada a complexidade de implicações no compartilhamento de dados abertos e no estabelecimento de uma cultura de Ciência Aberta, a disponibilidade de dados da pesquisa, um aspecto da Ciência Aberta, afeta a dinâmica do ciclo de pesquisa para geração de conhecimento científico.

Nesse contexto, trata-se do processo de publicação científica como uma rede de atores, cujos dados são actantes não humanos presentes na rede, trata-se igualmente da importância do compartilhamento e do reuso dos dados para a dinâmica da rede (BORGMAN, 2000; 2007; CURTY, 2016a, CURTY, 2016b; CURTY *et al.*, 2017; TENOPIR, 2011, 2015), ancora-se então a análise do cenário na Teoria Ator-Rede - TAR. Para tal identificam-se os atores envolvidos na produção científica, a partir dos principais conceitos do *corpus* da TAR, privilegiando a descrição dos atores, que revelam a rede e sua dinâmica. Por fim, aplica esses conceitos aos ambientes de pesquisa e sua relação com letramento informacional a fim de promover o reuso dos dados.

Ao aproximar os conceitos da TAR com a rede de comunicação científica, Estevão, Arns e Strauhs (2019, p. 11) destacam que:

Desta forma, sob a perspectiva da TAR (LATOURE, 2001), em que elementos humanos e não-humanos tecem uma determinada rede, pode-se inferir que os pesquisadores, as agências de fomento, as instituições de ensino, os laboratórios, os *bits*, os arquivos eletrônicos em quaisquer formatos, os sistemas de armazenamento da produção acadêmica, as bibliotecas acadêmicas, os mecanismos de buscas *online* para recuperação de informação e pesquisas, além dos próprios dados de pesquisa em si, constituem, dentre outros elementos, a rede de processo de investigação e geração de conhecimento científico.

Utilizar a perspectiva da TAR, portanto, permite que os elementos não humanos, como os dados da pesquisa, tenham a mesma valoração dentro da rede que os elementos humanos, explorando a noção de simetria generalizada.

Destacados os principais autores e autoras que tecem o pano de fundo para o arcabouço teórico que sustentará as discussões desta pesquisa, salienta-se que sequencialmente serão abordadas detalhadamente as questões da Comunicação Científica sob um viés sociotécnico, estabelecendo a Rede de Atores que cria dados de pesquisa exigindo, no entanto, noções de Letramento Informacional para o seu uso e reuso efetivos, dentro de uma abordagem

interpretativista.

De ótica qualitativa, portanto, as questões metodológicas se apropriam dos conceitos da Análise de Conteúdo de Bardin (2011) e da Revisão Sistemática de Guertin e Bernhardt (2005) para o estabelecimento de técnicas e procedimentos. Sequencialmente, apresenta-se a estrutura da Tese, com breve descrição dos Capítulos expressa graficamente na Figura 5.

As seções 1.6 e 1.7 apresentaram a abordagem metodológica e a fundamentação teórica, respectivamente. A abordagem metodológica na seção 1.6 deste capítulo introduziu um panorama das escolhas e condutas metodológicas da pesquisa, finalizando com o planejamento, sendo detalhadamente apresentados no capítulo 5. A fundamentação teórica inicial demonstrou um pano de fundo dos principais conceitos do tema de pesquisa, sendo desdobrados nos capítulos 2 e 3. Sequencialmente, apresenta-se a estrutura da tese, com a breve descrição dos capítulos e os números indicativos na Figura 5.

1.8 ESTRUTURA DA TESE

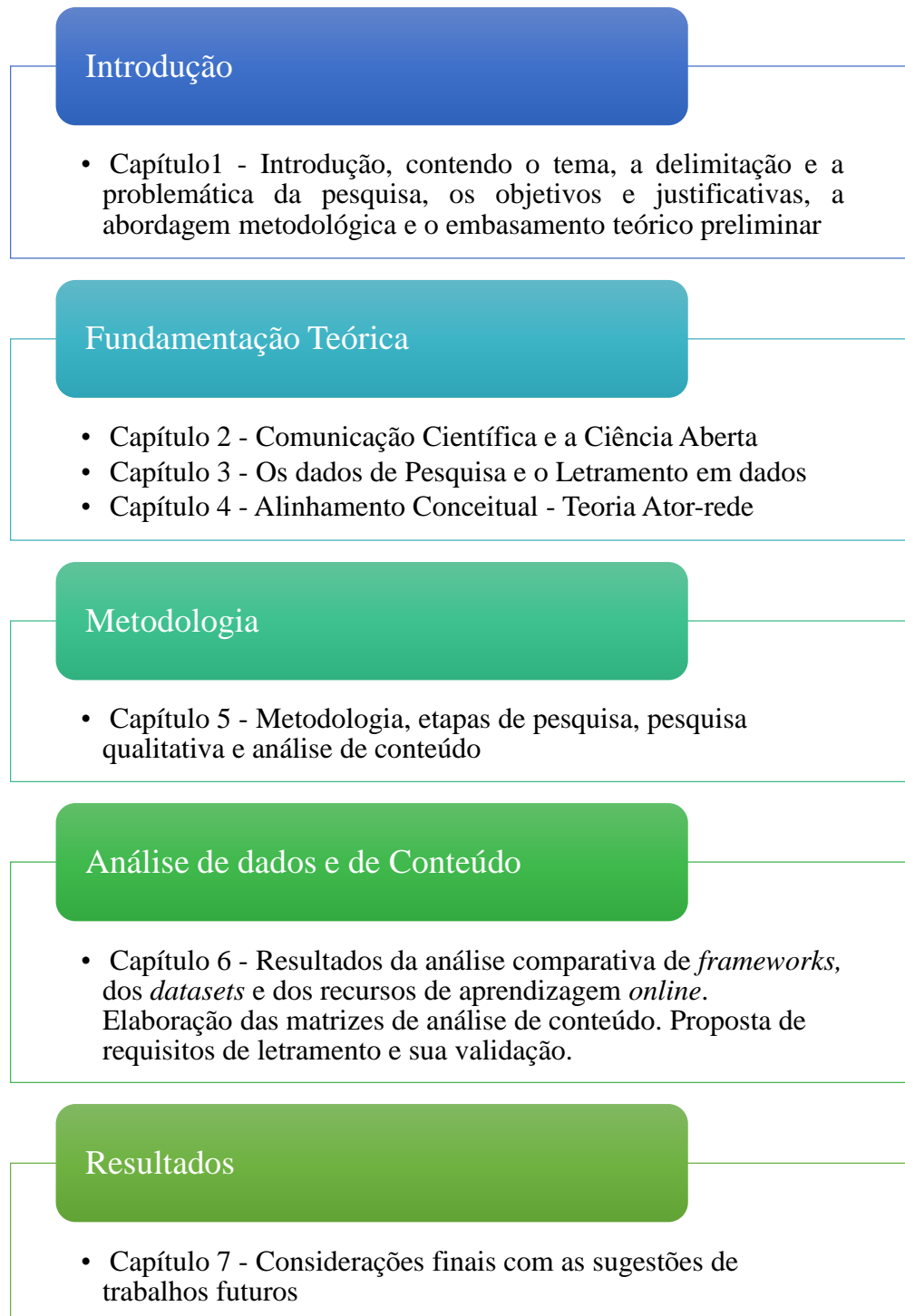
Esta tese está estruturada da seguinte forma: este Capítulo 1, introdutório, que aborda o tema, a delimitação e a problemática da pesquisa, seguido de seus objetivos e justificativas, a abordagem metodológica, o marco teórico e esta descrição de estrutura.

Um segundo bloco refere-se ao arcabouço conceitual, dividido nos Capítulos 2, 3 e 4. O segundo e terceiro Capítulos iniciam com as teorias de bases, seguidos da revisão bibliográfica de cada conceito abordado na pesquisa: a comunicação científica e o letramento informacional para o reúso dos dados de pesquisa. O Capítulo 4 destaca os conceitos selecionados como base dessa pesquisa, procurando evidenciar o prisma conceitual da pesquisa e realizando o alinhamento conceitual com a Teoria Ator-Rede.

O Capítulo 5 apresenta os procedimentos metodológicos, neste classifica-se, estrutura-se o planejamento e se detalha a execução da pesquisa.

O Capítulo 6 analisa os resultados dos dados, dos quadros conceituais de letramento (*frameworks*), dos recursos de aprendizagem *online* e realiza a discussão dos mesmos, Traz a proposta de requisitos de letramento e a sua validação. O Capítulo 7 registra as conclusões e aponta proposições de trabalhos futuros. A Figura 5 apresenta a proposta de estrutura desta tese.

Figura 5 - Estrutura da tese



Fonte: Autoria própria (2019).

A fundamentação teórica, composta pelos Capítulos 2, 3 e 4, no quais se abordam os *corpora* pesquisados da Comunicação Científica e da Ciência Aberta, dos Dados de Pesquisa e do Letramento em Dados, seguido pelo Alinhamento Conceitual que consolida a Teoria Ator-Rede com a revisão de literatura, são tratados sequencialmente.

2 COMUNICAÇÃO CIENTÍFICA: da Ciência Aberta à Publicação, breve cenário

O presente capítulo trata do atual contexto da pesquisa científica aberta, aliado ao modelo de disseminação da perspectiva de um sistema sociotécnico, além das demais funções da comunicação científica que extrapolam a divulgação das produções acadêmicas, à luz da Teoria Ator-Rede. Discute-se o aspecto do acesso aberto como via de publicidade dos resultados desse processo, estimulado e potencializado pela intensa exigência de leis nacionais e políticas institucionais internas das universidades e financiadoras de pesquisa em diversos países. O acesso aberto também está regulado nas políticas de revistas científicas nacionais, o que demonstra o nível de maturidade das revistas brasileiras em disponibilizar de forma gratuita o acesso ao conteúdo publicado, assunto igualmente discutido neste capítulo.

2.1 O PARADIGMA CIENTÍFICO DO SÉCULO XXI - A CIÊNCIA ABERTA E O eScience

Antes da metade do século XVII não era usual o compartilhamento dos resultados científicos (PROSSER, 2013). Havia um sentimento de que os pesquisadores eram guardiões do conhecimento e, portanto estes, eram relutantes em fornecer aquilo que poderia ser uma vantagem competitiva aos seus rivais. Ainda que os pesquisadores quisessem garantir os direitos da prioridade de suas “descobertas”, também não desejavam disseminar a informação (PROSSER, 2013, p. 40, tradução nossa). Um exemplo é a forma de publicação do descobrimento dos anéis de Saturno, por Galileu, em 1610. Seu trabalho foi publicado em forma de um anagrama de 37 letras em latim para garantir a prioridade da descoberta, porém, com a tentativa de que fosse de uma forma protegida (PROSSER, 2013).

Foi neste ambiente de segredos que surgiu o primeiro periódico científico. Esta era uma tentativa que buscava mudar o paradigma para o de que a publicação dos resultados é que garantiria a prioridade de uma determinada descoberta para o pesquisador (PROSSER, 2013). Com a contínua profissionalização da pesquisa, já na metade do século XIX, os cientistas publicavam seus achados regularmente e, às vezes, até de forma independente (PROSSER, 2013). A disponibilidade da publicação havia alterado o comportamento da atividade científica. Prosser (2013) também ressalta que a comunicação científica não está relacionada apenas com a publicação dos resultados acadêmicos, mas que é uma parte vital do próprio processo de pesquisa em si. Para esse autor, a forma como os pesquisadores disseminam os seus resultados tem sido profundamente alterada nas últimas décadas, desde o lançamento do primeiro

periódico científico há quatro séculos, alterando também o ambiente de pesquisa (PROSSER, 2013).

Não apenas o pesquisador, mas diversos atores estão envolvidos nesse processo. Além das transformações ocorridas na atuação dos pesquisadores (PROSSER, 2013), outros atores como os editores de periódicos (MCGRATH, 2013), os órgãos de financiamento (KILEY, 2013), os institutos de pesquisa (CARTER, 2013), as bibliotecas acadêmicas (BROWN, 2013) e os leitores (SCHONFELD, 2013) também vem passando por transformações.

Nessa perspectiva, Prosser (2013) defende que a atividade de publicação foi alterada como reflexo das mudanças no processo de pesquisa, porém o processo de pesquisa em si também foi alterado como resultado dos desenvolvimentos da comunicação científica. Ainda que o processo tenha sofrido transformações essencialmente em decorrência dos avanços tecnológicos e apesar das possibilidades das novas tecnologias em termos de formatos de publicações e canais de distribuição ou modelos de negócios, os propósitos ou finalidades da comunicação científica têm permanecido, notavelmente, estáveis (BORGMAN, 2007).

A Ciência Aberta, neste desenvolvimento, representa uma nova abordagem para o processo científico com base no trabalho cooperativo e em novas formas de difusão de conhecimentos por meio de tecnologias digitais e das novas ferramentas colaborativas (EUROPEAN COMMISSION, 2016c). A Organização para Cooperação e Desenvolvimento Econômico – OCDE (ORGANIZATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT, 2015, p. 7, tradução nossa) conceitua Ciência Aberta como uma forma de "permitir o acesso aos resultados primários de pesquisa financiada com recursos públicos – publicações e os dados de pesquisa – disponíveis ao público em formato digital sem ou com restrição mínima". Porém, para a própria OCDE (2015), é mais do que isso, Ciência Aberta é frequentemente conceituada como um termo abrangente que envolve vários movimentos tendentes a suprimir as barreiras para compartilhar qualquer tipo de resultados, de recursos ou de dados da pesquisa, de métodos ou de ferramentas, em qualquer fase do processo de investigação científica (ORGANIZATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT, 2015).

De acordo com a taxonomia do Facilitate Open Science Training For European Research, (2015, *online*)¹⁹, a Ciência Aberta é a prática da ciência, de tal forma, que outros possam colaborar e contribuir; em que dados de pesquisa, notas de laboratório e outros

¹⁹ FACILITATE OPEN SCIENCE TRAINING FOR EUROPEAN RESEARCH. **Open Science Definition.** 2015. Disponível em: <https://www.fosteropenscience.eu/foster-taxonomy/open-science-definition>. Acesso em: 02. Jul. 2019.

processos de pesquisa estejam disponíveis gratuitamente de forma a permitir a reutilização, a redistribuição e a reprodução da pesquisa e seus dados e métodos. A Ciência Aberta representa uma abordagem à pesquisa que é colaborativa, transparente e acessível, considerando aspectos tanto culturais e de conhecimento, quanto de tecnologias e de serviços em um movimento que busca tornar os produtos e processos científicos acessíveis e reutilizáveis por todos (GITBOOK BOT *et al.*, 2018). A perspectiva da Ciência Aberta transcende o acesso aos resultados de pesquisas, visto que uma ampla gama de atividades está sob a égide da Ciência Aberta – e nas suas fronteiras, incluindo as Publicações de Acesso Aberto, os Dados Abertos, as Notas abertas, as Revisão por Pares Aberta e a Educação Aberta (EUROPEAN COMMISSION, 2017).

Fecher e Friesike (2014) apresentam algumas categorias de abordagem do termo “Ciência Aberta” de acordo com uma revisão de literatura da área. A noção geral de Ciência Aberta engloba uma variedade de premissas, porém, sobre um mesmo constructo: o futuro da criação e da disseminação do conhecimento. Fecher e Friesike (2014) apresentam cinco abordagens sobre Ciência Aberta, que chamam de escola de pensamentos, baseadas na percepção de vários cientistas e pesquisadores que publicaram sobre o assunto. A conceituação do termo depende, portanto, do ponto de vista das partes interessadas que serão diretamente afetadas no ambiente de pesquisa científica (FECHER; FRIESIKE, 2014).

As abordagens identificadas por Fecher e Friesike (2014) sobre a Ciência Aberta, considerando a diversidade das partes envolvidas são: (a) escola democrática, que preconiza a igualdade de acesso; (b) escola pragmática, que evoca a pesquisa colaborativa; (c) escola de infraestrutura, que está relacionada com arquitetura tecnológica que viabilize ferramentas colaborativas; (d) a escola de publicidade, que diz respeito a acessibilidade na criação do conhecimento; e (e) a escola de métricas, pautada nas métricas alternativas de impacto. Neste estudo, adota-se os preceitos das escolas democrática e a pragmática pois são as que incluem os dados abertos na sua abordagem, conforme citado na Seção 1.2 – Capítulo 1.

O Quadro 1 apresenta as principais prerrogativas de cada abordagem, os grupos envolvidos, assim como a principal argumentação central, os métodos e as ferramentas para viabilizar cada aspecto da Ciência Aberta categorizados por Fecher e Friesike (2014).

Quadro 1 - Comparativo das abordagens da Ciência Aberta

Abordagem	Premissa Central	Partes envolvidas	Argumentação central	Ferramentas e Métodos
Democrática	O acesso ao conhecimento não é igualmente distribuído	Cientistas, políticos e cidadãos	Tornar o conhecimento disponível gratuitamente a qualquer um	Acesso aberto , direitos autorais, dados abertos , <i>software</i> aberto
Pragmática	Criação do conhecimento poderia ser mais eficiente se os cientistas trabalhassem em colaboração	Cientistas	Abertura do processo de criação do conhecimento	Sabedoria das multidões, efeitos de rede, dados abertos , <i>software</i> aberto
Infraestrutura	Pesquisa eficiente depende de ferramentas e aplicações disponíveis	Cientistas e Provedores de plataformas	Criação de plataformas abertas, ferramentas e serviços para cientistas	Ferramentas e plataformas colaborativas
Pública	Ciência precisa se tornar acessível ao público	Cientistas e cidadãos	Tornar a ciência acessível aos cidadãos	Ciência cidadã, <i>Blogging</i> científico
Métricas	As publicações científicas precisam de métricas alternativas de impacto	Cientistas e políticos	Desenvolver um sistema alternativo de métricas de fator de impacto	Altimetria, revisão por pares, citação e fatores de impacto

Fonte: Adaptado de Fecher e Friesike (2014, p. 20, tradução nossa).

Diante das concepções mais amplas de Fecher e Friesike (2014), a Ciência Aberta pode ser considerada um importante mecanismo para a viabilidade da pesquisa científica contemporânea, que utiliza intensivamente dados digitais de pesquisa (SAYÃO; SALES, 2016). Essa abordagem que considera o uso intensivo de dados na prática científica se denomina *eScience*.

Cunhado por John Taylor (2001), que foi diretor-geral do Conselho de Pesquisa do Office of Science and Technology (OST) do Reino Unido o termo *eScience* foi difundido, posteriormente, por James Nicholas Gray, inicialmente por meio de palestras e de discussões com pares. A perspectiva de pesquisa científica de Taylor (2001) para o século XXI possibilitaria aos pesquisadores lidar de forma mais eficaz com o grande volume de dados gerados.

Existem diferenças na forma escrita de *e-Science*, como *eScience*, *EScience*, *escience* e *Science*, e a literatura também aponta para termos sinônimos como Ciência Orientada a Dados (*Data-Driven Science*), Pesquisa Eletrônica (*e-Research*), Ciber Infraestrutura (*Cyberinfrastructure*), Ciência Melhorada (*Enhance Science*), Ciber Ciência (*Cyberscience*) e Infraestrutura Eletrônica (*eInfrastructure*) (CESAR JUNIOR, 2011, p.7; MEDEIROS; CAREGNATO, 2012, p. 315) bem como se aponta para um quarto paradigma da ciência (*fourth*

paradigm of science) (HEY; TREFETHEN, 2003).

Portanto, o termo *eScience* se configura como uma prática de pesquisa científica intensiva na geração de dados de pesquisa e também no seu uso, no armazenamento, no processamento, na análise e no compartilhamento destes (HEY; TANSLEY; TOLLE, 2009). Em 2014, o Institute of Electrical and Electronics Engineers (IEEE) resumiu a conceituação, constatando que o "*eScience* promove a inovação em pesquisas colaborativas, computacionais ou intensivas em dados em todas as disciplinas ao longo do ciclo de vida da pesquisa" (INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS, 2014, *online* – tradução nossa). Para Medeiros e Caregnato (2012), a *eScience* modifica a forma que os cientistas trabalham, desde os tipos de problemas abordados quanto as ferramentas que utilizam, alterando profundamente a natureza da documentação e da publicação resultantes da sua pesquisa (MEDEIROS; CAREGNATO, 2012).

Sendo assim, a prática da ciência é realizada de tal forma que outros possam colaborar e contribuir, sempre que os dados da pesquisa, notas de laboratório e outros processos de pesquisa estejam disponíveis gratuitamente e de forma facilitada, em termos que permitam a reutilização, a redistribuição e a reprodução da pesquisa e seus dados. Medeiros e Caregnato (2012) identificam que além da colaboração e da cooperação, a *eScience* possibilita a interdisciplinaridade. Profundas alterações, portanto, estão sendo apontadas no processo de comunicação científica nas últimas décadas enquanto atividade sociotécnica (BORGMAN, 2007; BORGMAN; FURNER, 2012; JUBB, 2013).

2.2 COMUNICAÇÃO CIENTÍFICA COMO UM SISTEMA SOCIOTÉCNICO

Borgman (2007) atesta que a comunicação é um sistema sociotécnico rico e complexo que foi sendo formado ao longo de séculos (BORGMAN, 2007, p. 48). Os sistemas sociotécnicos têm como característica uma complexa inter-relação de atores, sejam estas pessoas, instituições, artefatos ou ambientes virtuais, elevando a complexidade e exigindo o pensamento sistêmico para lidar com a interdependência desses elementos (CALLON; LAW, 1989; LINSINGEN; BAZZO; PEREIRA, 2003).

Callon, Law e Rip (1986) concordam que tanto a ciência como a tecnologia são forças poderosas na sociedade moderna industrializada, entretanto, a forma como sua força é criada e desenvolvida é obscura. Segundo Callon, Law e Rip (1986, p. 4, tradução nossa), "a ideia que há um método científico especial, um domínio onde a verdade prospera em detrimento do poder,

é um mito”. O desenvolvimento da Ciência pode ocorrer de muitas formas e há um pensamento positivista que rodeia todo o processo. Linsingen, Bazzo e Pereira (2003) explicam as origens dessa concepção essencialista e triunfalista da Ciência e da tecnologia (C&T), nascida com o fim da segunda grande guerra, em um ambiente de otimismo e esperança nas possibilidades da C&T.

O modelo tradicional de C&T é amparado na ideia de relação diretamente proporcional de que quanto mais Ciência, mais tecnologia, mais riqueza e conseqüente maior o bem estar-social, pensamento denominado de determinismo²⁰ científico e tecnológico, conceito que o campo de estudo CTS está há décadas ressignificando e revelando as influências sociais implícitas e explícitas na atividade científica, demonstrando que a ciência e a comunicação dos seus resultados são um processo sociotécnico (BORGMAN, 2007; CUTCLIFFE, 2003).

Callon, Law e Rip (1986, p. 3, tradução nossa) defendem que “um entendimento da dinâmica da ciência somente é possível quando a força da ciência no dia a dia da sociedade é considerada”. Esse entendimento correto da mudança social e científica requer o abandono da dicotomia entre ciência e política, sendo que a compreensão dessas transformações pressupõe que os atores sejam seguidos, não somente nos palácios (quando fazem política), mas também nos laboratórios (quando fazem ciência) (CALLON; LAW; RIP, 1986). Sendo assim, “eles (cientistas, pesquisadores) não estão apenas praticando ciência, eles estão também praticando política, economia e sociologia” (CALLON; LAW; RIP, 1986, p. 10). Cutcliffe (2003, p. 18), escrevendo sobre a C&T, explica que:

[...]a Ciência e Tecnologia são projetos complexos que se dão em contextos históricos e culturais específicos, nos quais valores culturais, políticos e econômicos nos ajudam a configurar os processos tecnocientíficos, os quais, por sua vez, afetam aos próprios valores e à sociedade que os sustentam.

A C&T é vista, então, como local de ação estratégica no qual existem relações sociais e científicas que trabalham para produzir, em um único movimento, o novo conhecimento e as novas relações sociais (CALLON; LAW; RIP, 1986). Para isso, as atividades práticas do cientista e as suas inscrições são muito importantes. O cientista busca “fazer uso de vários tipos de recursos para colocá-los juntos para se beneficiar dos resultados” (CALLON; LAW; RIP, 1986, p. 9, tradução nossa).

Os elementos com os quais os cientistas constroem sua versão de mundo passam pela

²⁰ Ideia de que a tecnologia influencia a sociedade em uma via de mão única, sem ser influenciada pela mesma (FEENBERG, 2010).

força manual, das máquinas, do dinheiro, dos discursos, dos textos, das reações químicas, das teorias, entre outros. “Cientistas atribuem especial importância para inscrições: gráficos, *preprints* e artigos científicos publicados, etc.” (CALLON; LAW; RIP, 1986, p. 10, tradução nossa). O texto, necessário na ciência, é uma ferramenta política central para o cientista, segundo Callon, Law e Ripp (1986, p. 10-11, tradução nossa):

[...] no texto um mundo estruturado é construído [...]. A força de tais textos também reside no fato de que eles contêm ligações com outros textos (citações e referências bibliográficas), trabalho e instituições [...] Em resumo, os textos tornam possível a construção de ligações entre entidades existentes e a formação de novas entidades e, se persuasivo, assim constitui um método importante para tentar controlar o ambiente.

Não é intenção aqui reduzir as atividades científicas e tecnológicas à propagação da inscrição literária, no entanto, considera-se que as inscrições são uma característica central da Ciência, pois o texto revela a estratégia de construção de mundo dos seus autores (CALLON; LAW; RIP, 1986). A comunicação da produção científica é uma etapa da pesquisa científica e está centrada na divulgação dos resultados das pesquisas, sejam por grupos de pesquisadores ou de indivíduos discentes e docentes, não apenas, mas sobretudo dos programas de pós-graduação do país ou de entidades financiadas com recursos públicos e privados. A comunicação científica é, portanto, composta de uma variedade de atividades que desempenham funções relacionadas à publicação e à disseminação dos resultados de pesquisa.

É a partir da descrição dos atores que se torna possível identificar a constituição e a dinâmica das redes. A proposta metodológica de Latour (2000; 2001) para o estudo da construção dos fatos se dá, de forma simples, como denominado por este, e conforme já citado, a partir do melhor de todos os guias: seguir os próprios cientistas. Para Latour (1998, p. 110), seguir os atores ou os cientistas significa descrever o que eles fazem, tanto quando eles tentam transformar a sociedade, quanto quando eles buscam construir o conhecimento científico ou o sistema tecnológico. Para esse autor, a sociedade é construída, mas não socialmente, antes “os humanos estenderam suas relações sociais a outros atuantes com os quais trocaram inúmeras propriedades, formando coletivos” (LATOUR, 2001, p. 227). Atores são as entidades que fazem coisas, que atuam, portanto, só são consideradas existentes se deixam ‘provas’ ou ‘rastros’ da sua agência (LATOUR, 2006, p. 346), constituindo redes, a partir da agência dos atores, assunto tratado sequencialmente.

2.3 DO PROCESSO DE COMUNICAÇÃO CIENTÍFICA À REDE DE ATORES

Os modelos tradicionais de comunicação científica, segundo Sayão e Sales (2012) abrangem a disponibilidade dos resultados de pesquisas, como artigos, teses e dissertações, após o pesquisador percorrer todas as etapas envolvidas, desde a pesquisa para elaboração do documento, as fases de avaliação por pares, a publicação, o tratamento, a disseminação, o acesso até a assimilação e o uso desses resultados para criação de novas pesquisas e, conseqüentemente, novos documentos. Esse contexto remete a um coletivo de actantes (atores) que demonstram sua presença na rede pela agência que exercem. Infere-se, a partir de Latour (2000; 2001), que ao seguir os atores dos modelos tradicionais de comunicação científica (pesquisadores, universidades, pares da comunidade científica, agências de financiamento, revistas, editores científicos e seus sistemas de busca *online*, a própria sociedade), é possível ver como estes exercem agência na dinâmica da rede de comunicação científica.

Uma rede, para Latour (1986), na perspectiva da Teoria Ator-Rede – TAR, suporte metodológico subliminar tácito deste estudo, refere-se a fluxos, a circulações, a alianças, a movimentos, não se remetendo a uma entidade fixa, mas às conexões ou às interações que transformam os recursos dispersos em uma rede.

Aproximando a comunicação científica da perspectiva de rede de Latour (2000), pode-se encontrar em autores como Roosendaal e Geurts (1997), uma análise da transformação da comunicação científica que deixa de ser uma cadeia de informações lineares e se transforma em uma rede de comunicação interativa. Essa evolução se dá como resposta às mudanças concomitantes na pesquisa científica e na educação. Para esses autores, essa transformação é de natureza estrutural, necessária para suportar as crescentes demandas sociais pelo crescimento e pela gestão do conhecimento. As publicações científicas são consideradas as principais operadoras para uma "troca heterogênea de conhecimento que é seminal para o progresso da ciência" (ROOSENDAAL; GEURTS, 1997, p. 3, tradução nossa) Esses autores defendem, portanto, o aumento da eficiência e da eficácia da comunicação científica, argumentando para tal que a sua configuração deve atrair tanto a comunidade de pesquisa quanto a sociedade em geral.

Ao abordar o papel de comunicação e sua organização em relação a todo o processo de pesquisa, remete-se à questão da complexidade da aparente abundância de informação científica. O paradoxo é que a quantidade excessiva de produção do trabalho científico original não pode atrasar o progresso científico, mas o faz, já que o excesso, e não a escassez de informações e dados, tem sido um problema para o pesquisador pela falta de tempo de pesquisa

para explorar o excesso de recursos informacionais com foco na melhoria da eficácia e da eficiência do conhecimento científico (ROOSENDAAAL; GEURTS, 1997). Para lidar com essa complexidade, esses autores defendem novos métodos de comunicação científica que permitam o avanço da ciência em um formato de uma "rede" ao invés de um processo linear da cadeia de informações, tendo a tecnologia como uma importante força motriz.

Redes devem ser, e são, construídas por atores que têm ação, interferência, ou na terminologia de Latour (2006, p. 340), tem “agência”, na perspectiva da Teoria Ator-Rede – TAR. A TAR não faz parte de uma rede pronta para ser descrita, mas de uma rede que se constrói a partir da descrição de atores que exercem ação (LATOURE, 2006). A TAR não é, portanto, uma configuração fechada pré-estruturada de uma rede, antes é uma forma de descrever a agência de atores que constitui a rede. As estruturas são engendradas pelos movimentos e pelas relações entre os atores, e não os atores se encaixam em um sistema previamente estruturado. A TAR também não considera uma explicação para os fatos, não é um quadro explicativo (LATOURE, 2006). A materialidade relacional da Teoria Ator Rede (TAR) é a ideia de que tudo é constituído em interação, e que nada tem existência definida fora destas interações (LATOURE, 2006). São as interações dos atores que formam um conjunto relativamente estável de entidades associadas que alcançam certa durabilidade. Isso os torna “actantes”, ou seja, só pode ser considerado um ator aquele (ou o que) exerce ação nas relações com outros elementos da rede (LATOURE, 2006, , p. 340). O ator é aquele que promove a aliança, ou seja, quem ou o que realiza o aliciamento dos demais entes para um objetivo. O conceito de ator é central na TAR, pois é por meio da sua força em mobilizar outros que a rede funciona.

Latour (2006) defende que o nome da teoria deveria trazer na sua composição o termo *worknet* (trabalho da rede) em vez de *network* (rede de trabalho), dada à importância na noção de ação, de movimento, de inter-relação. Se o elemento não faz a sua presença ser percebida pela sua influência na estrutura da rede, então se pressupõe que para aquela rede, este elemento não existe (LATOURE, 2006; LAW, 1993). Sendo assim, a extensão da rede é definida pela presença de atores que são capazes de se fazerem percebidos de uma maneira individual (LAW, 1993). Considerando este escopo, foram identificados alguns atores da rede de geração de conhecimento científico e como a disponibilidade dos dados altera a dinâmica desta rede, como por exemplo, a localização de conjuntos de dados disponíveis sobre um determinado objeto de pesquisa, da mesma forma que se evidencia o estado da prática por meio de levantamentos bibliográficos.

Desta forma, dentro de uma abordagem TAR (LATOUR, 2001), pode-se inferir que sob uma perspectiva simétrica, atores serão os pesquisadores, as instituições de fomento e de ensino, os arquivos eletrônicos, os sistemas de buscas *online* e os resultados da produção acadêmica, as bibliotecas e os próprios dados de pesquisa. Descrever esses atores é construir a rede, e é isso que trata a seção seguinte.

2.4 DESCRIÇÃO DOS ATORES NA REDE DE COMUNICAÇÃO CIENTÍFICA

Os dados não são o único ator dentre os elementos não humanos. Analisando sob a perspectiva da simetria generalizada²¹, diversos outros actantes não humanos e humanos interagem na dinâmica da rede da comunicação científica. Um elemento não humano são os Repositórios Institucionais de Dados. A partir do movimento pelo acesso aberto às políticas de acesso, os repositórios institucionais alteraram a forma e, principalmente, o custo com que os resultados de pesquisa são disponibilizados (RODRIGUES, 2008). Com a constante intensidade em geração e uso dos dados, os repositórios institucionais de dados de pesquisa e suas políticas de gestão têm propiciado o suporte para a que os arquivos digitais de dados tenham um papel relevante, alterando a própria forma de construção da ciência (HEY; TANSLEY; TOLLE, 2009).

Para entender a forma de agência dos não humanos, Latour (2001) apresenta o conceito de mediação técnica. Latour (2001) denominou a atuação não humana como uma mediação técnica. A compreensão da ideia de mediação técnica é muito importante para o entendimento de como se dá a agência dos chamados elementos não humanos. A primeira forma da mediação técnica são os pressupostos da agência dos não humanos na medida em que estes interferem na ação humana, atuando como mediadores e realizando a translação de objetivos.

A mediação técnica pode ser usada como uma lente ao observar as interações na comunicação científica. Na medida em que um elemento não humano, como um artigo ou um conjunto de dados, divulga e constrói fatos científicos de um pesquisador que não está ali presencialmente, esse actante não humano utiliza-se então de suas propriedades técnicas para comunicar aquilo que o cientista deseja. Os artefatos “falam” aos leitores, trocando propriedades com seus idealizadores (LATOUR, 2001, p. 216). É uma coisa que age

²¹ A simetria generalizada é o pressuposto que atribui o mesmo valor de análise de relações de elementos não humanos e os humanos em uma dada rede, sem privilégios ou seleções (LATOUR, 1992) Em última instância, a ideia de simetria generalizada é de que a que a sociedade e a natureza seriam coproduzidas (LATOUR, 1992).

confiantemente como legado. Em última análise, os registros ou os dados da pesquisa não são feitos apenas por *bits*, tendo sua leitura intermediada por *softwares* e máquinas, mas estão repletos de cientistas que misturam suas pesquisas e seus perfis aos programas de computador, aos caracteres, aos algoritmos e aos transistores, em uma relação heterogênea entre humanos, máquinas e inscrições.

Essa ação colaborativa, segundo Latour (2001), é a mediação técnica acontecendo como partes de uma composição, em que a ação é uma das propriedades das entidades associadas transpondo a fronteira entre signos e coisas que provoca o deslocamento de significado, em que a sociedade (humanos) e a matéria (não humanos) trocam propriedades (LATOURE, 2001). Ou seja, a ação não é uma propriedade de humanos, mas é o resultado de uma associação de diversos atuantes. Os objetivos são redefinidos por associações em conjunto com atuantes não humanos, e a ação é uma propriedade ou um resultado da associação inteira, incluindo os não humanos e não apenas dos atuantes chamados humanos (LATOURE, 2001). O alinhamento com estas diretrizes permite inferir como os dados de pesquisa exercem agência na rede de geração do conhecimento científico, ou seja, um elemento não humano que, em associação com os humanos e com máquinas, atuam de forma colaborativa na produção e na comunicação científica.

2.4.1 Os Dados de Pesquisa como Ator Não Humano da Rede

A partir da proeminência dos dados evidenciada na literatura, infere-se que os dados de pesquisa constituem o elemento central que subsidia os resultados da pesquisa, pois são os dados coletados e utilizados para análise que serão, posteriormente, tratados e condensados em forma de resultados divulgados nos textos científicos. Se as inscrições são os alicerces dos textos científicos, como afirma Latour (2000), logo, os dados de pesquisa fazem parte da sustentação dos resultados da própria pesquisa. Utilizar a abordagem TAR permite que os elementos não humanos, como os dados da pesquisa, tenham a mesma valoração dentro da rede que os elementos humanos – simetria generalizada.

A simetria generalizada é um conceito da Sociologia da Ciência que pressupõe dar o mesmo tipo de explicação para crenças verdadeiras e falsas, já que as verdadeiras eram consideradas realidades e as falsas como explicações psicológicas ou sociais (LAW, 1993). Latour (2000) indica que essa simetria é na verdade uma assimetria, pois pressupõe colocar todos o peso da explicação das causas naturais na sociedade, desconsiderando o não social, as

coisas naturais e os processos e como estes interferem nos próprios arranjos sociais. Considera-se, portanto, que os dados de pesquisa são um ator-rede, pois sua presença é marcada por alterações desdobradas e percebidas na rede, pela mudança que provoca na dinâmica de produção do conhecimento. Os dados, quando disponíveis, alteram a forma como os pesquisadores podem produzir análises, inferências, comparações, resultando no próprio conhecimento científico.

Infere-se que o pesquisador é quem absorverá mais uma atividade, que é a publicação dos dados da pesquisa, que antes ficavam subutilizados no modelo tradicional. Esta atividade compreende todas as etapas do ciclo de vida dos dados. Entretanto, as implicações desta mudança são diretamente benéficas à própria atividade de pesquisa. Além dos sistemas tradicionais de buscas, que requerem pagamento para acesso, os recursos abertos, viabilizados pela retroalimentação da atividade científica, ficam disponíveis ao pesquisador, ampliando seu escopo de fontes de informação de forma gratuita. O acesso aos repositórios institucionais que trazem os resultados de outras pesquisas e aos repositórios de dados de outros pesquisadores permite a democratização da produção científica e em última instância, beneficia a sociedade como um todo.

Diante da proeminência dos dados de pesquisa evidencia-se como a disponibilidade dos mesmos interfere na geração do conhecimento e de comunicação científica. Na medida em que o pesquisador pode valer-se de novos insumos de informação, ou seja, os conjuntos de dados de pesquisa disponibilizados por outros pesquisadores para análise, comparação, aplicação e reúso desses dados na sua própria pesquisa, usando os repositórios de dados para a busca destes conjuntos de dados, ele também precisará, posteriormente, disponibilizar os dados que produziu nos mesmos repositórios, em um processo interativo.

2.4.2 As Interações dos Dados com os demais Atores da Rede de Comunicação Científica

Observando-se o cenário tradicional, quando apenas um levantamento bibliográfico foi realizado, se comparado ao contexto estimulado pela Ciência Aberta, infere-se que a possibilidade da busca por dados pode alterar a metodologia de pesquisa do investigador. Os dados produzidos como parte das pesquisas têm uma ampla variedade de formatos, desde estatísticas e resultados experimentais, até registros e transcrições de entrevistas (BORGMAN, 2012). Há uma expressiva quantidade desses dados disponíveis nos repositórios de dados. Contudo, a disponibilidade dos dados de pesquisa requer de outros actantes as suas atuações,

como por exemplo, os próprios repositórios de dados, que têm requisitos distintos dos repositórios institucionais.

Um ator, na concepção da TAR, precisa efetivamente arregimentar outros atores para que a sua ação seja ressonante na dinâmica da rede. Sem as características inerentes aos repositórios de dados, os dados não podem ser acessíveis e, conseqüentemente, reutilizáveis. Neste sentido, a ideia de mediação técnica da TAR pode ser entendida, pois a viabilidade técnica é o que mediará o acesso aos dados. Esta condição inerente ao elemento não humano condiciona a atividade humana para um determinado objetivo, o que nas palavras de Latour (2001, p. 218) é explicado por atores “trocando propriedades”.

Exemplificando, entidades como Agências de Financiamento (KILEY, 2013) e Instituições de Pesquisa, como universidades e centros de pesquisa (CARTER, 2013) sem dúvida têm exercido agência relevante no estabelecimento de políticas institucionais que requerem a disponibilidade de dados das pesquisas que financiam.

Os Editores dos periódicos científicos exercem, por seu turno, um papel de controle na rede de comunicação científica, considerando o atual modelo de disseminação da informação pautado pela produção, venda e distribuição de publicações (BORGMAN, 2007). Tenopir e King (2009) identificaram que, dentre os 225 mil títulos de jornais do mundo, 63 mil são periódicos científicos, dos quais aproximadamente 25 mil são os títulos principais. Porém, as mudanças no contexto social que evidenciam maior cultura de abertura e transparência do investimento público, estão ameaçando os modelos de negócios dos periódicos e a estabilidade desse sistema, não só pela prática de preços monopolísticos, mas também pelo crescimento do movimento do acesso aberto. Mcgrath (2013) relata que o acesso aberto tem um impacto distinto de acordo com o campo do conhecimento. Ainda que não se tenha respondido até o momento à questão de como criar modelos de negócios viáveis sem envolver o pagamento por conteúdo nem de editores ou de autores, a tendência é que cada vez mais, em ritmos distintos em cada disciplina, haja uma crescente disponibilidade de conteúdo sem custo de acesso (MCGRATH, 2013).

Assim como muitos artigos estão sendo enviados para repositórios e aumentando o número de publicações que não estão mais sujeitas a embargo, a existência do periódico como artefato para aglutinar artigos está sendo repensada. Mesmo que ocorra em um tempo de escala muito variável entre as disciplinas, a maioria dos artigos estará provavelmente disponível gratuitamente entre 10 a 15 anos nos repositórios de acesso aberto (MCGRATH, 2013). Assim as universidades e outros organismos de pesquisa podem novamente se tornar as principais fontes de publicação e distribuição (MCGRATH, 2013).

Os colaboradores do periódico científico (que são os autores, revisores e editores) oferecem seu serviço sem remuneração, porém, o editor as vezes recebe um pequeno honorário (MCGRATH, 2013). Os leitores, que muitas vezes também exercem o papel de colaboradores, leem as publicações sem custo, enquanto a biblioteca acadêmica é quem paga ao publicador pelo acesso ao conteúdo. A biblioteca acadêmica é, portanto, um importante actante na retroalimentação da comunicação científica, que intermedia a disponibilidade dos resultados de pesquisa como fontes de informação para novas produções científicas (BROWN, 2013).

Nesse contexto, a indústria do conhecimento é caracterizada como uma rede complexa de diferentes atores que são os produtores de conteúdo (autores, revisores e editores), organizados em diferentes tipos de instituições de pesquisa, de bibliotecas, de publicadores, de intermediários como os agentes de assinaturas e ainda os desenvolvedores de *softwares*, que viabilizam, assim, a indústria da comunicação científica. Esses são os mais recorrentes na literatura que exercem agência e cujas relações constituem a rede de comunicação científica (BORGMAN, 2007; BROWN, 2013; CARTER, 2013; KILEY, 2013, MCGRATH, 2013; PROSSER, 2013; SCHONFELD, 2013). Ainda que sejam vários, a coordenação dos interesses passa por vínculo em comum a todos eles, o ponto de passagem obrigatório (PPO), para viabilizar os objetivos individuais e coletivos.

2.4.3 Dados de Pesquisa como Ponto de Passagem Obrigatório – PPO

Sob o princípio da simetria generalizada, todos esses elementos heterogêneos que compõem a rede são considerados como componentes de um sistema (LAW, 1993) usando como a analogia os *systems builders*²² de Hughes (1993). É como dar voz a esses elementos para que eles façam suas próprias explicações. Porém, cada elemento tem um interesse e compreender interesses tão variados, conduzindo-os para que todos trabalhem em conjunto e em acordo (SISMONDO, 2010), de forma simétrica, é o processo denominado dentro da TAR de sociologia da tradução (CALLON, 1986).

O conceito de tradução é um importante pilar da TAR. Este é compreendido como o alistamento de atores em busca de um objetivo em comum, o que se convencionou denominar **ponto de passagem obrigatório** – PPO (CALLON, 1986). O ponto de passagem obrigatório é o vínculo das relações na rede (CALLON, 1986). A noção de tradução compreende, ainda, um

²² Construtores de sistemas.

processo de contínua mudança de objetivos, forças e interesses que ocorrem nos diversos estágios das relações entre natureza e sociedade (CALLON, 1986).

Diante das relações na rede de comunicação científica, a disponibilidade dos dados de pesquisa constitui o PPO, já que o interesse e os objetivos de diversos atores estão orientados a esta disponibilidade (Ciência Aberta, aspectos sociais, aspectos regulatórios, governança, eficiência dos recursos públicos, acessibilidade dos dados, conteúdo dos dados, aplicabilidade dos dados, reúso dos dados). Quando há a disponibilidade dos dados, eles passam a ter sua própria voz. Eles não estão mais limitados às análises e às interpretações que aprouve ao cientista explorar. Eles são um ator que tem suas próprias versões, sua própria “voz” pode ser ouvida diretamente pelo aceso na íntegra das evidências que eles carregam.

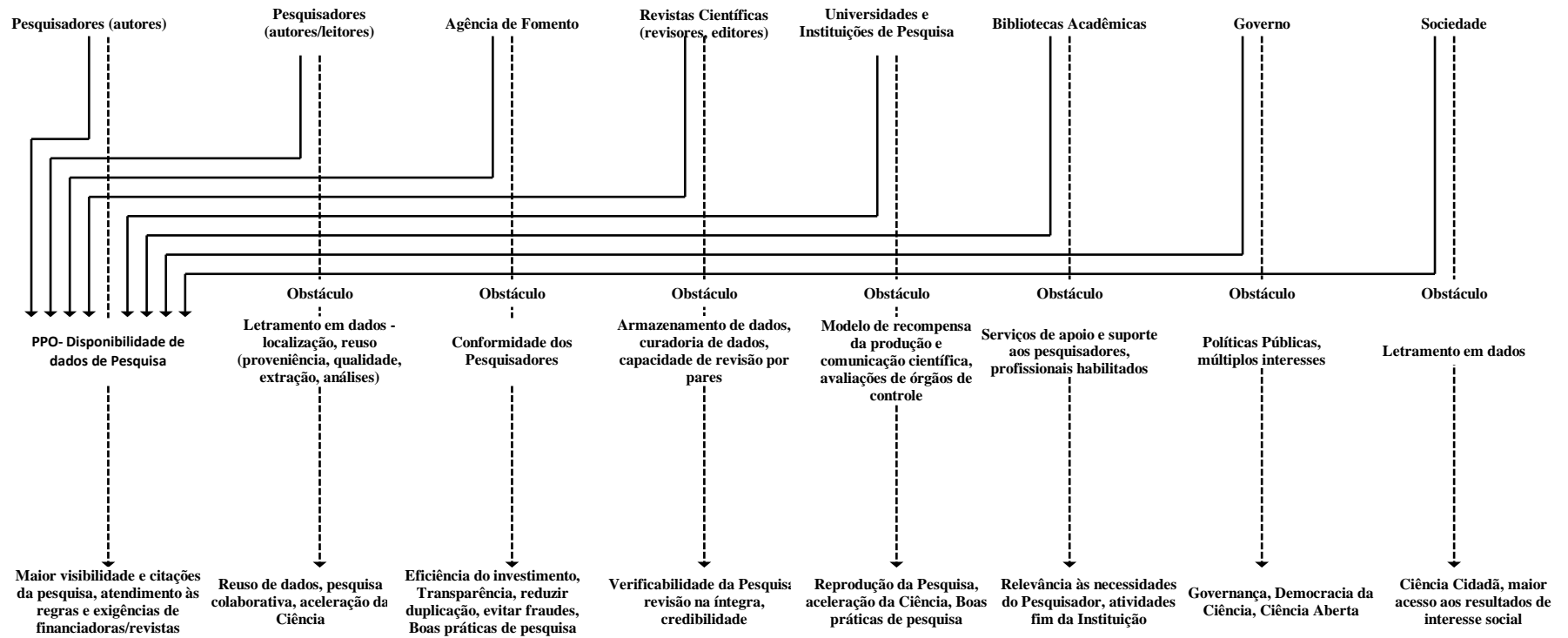
Além dos próprios autores e leitores, no entanto, é preciso olhar também a comunidade acadêmica de forma mais ampla, considerando os pesquisadores que também atuam como revisores e colaboradores na produção dos dados e dos resultados de pesquisa. Esses resultados, por sua vez, ficam depositados em repositórios institucionais e de dados, a partir de políticas das agências de financiamento e dos objetivos das instituições de pesquisa, tendo a sua publicação e o acesso viabilizados pelos editores e pelas bibliotecas acadêmicas.

Roosendaal e Geurts (1997) já afirmavam que a pesquisa estratégica requer novas formas de gestão do conhecimento individual do pesquisador e também do Instituto de pesquisa. Nesse contexto, eles formularam questões sobre o papel futuro das publicações e dos tipos de publicação requeridos pela prática da pesquisa, instrumentos de explicitação do conhecimento criado. Eles percebiam também um avanço significativo na mudança do foco (processo de tradução – contínuo alistamento de atores para objetivos comuns). Como um exemplo de que a atual produção científica precisa de formas alternativas de certificação da informação dos resultados de pesquisa, esses autores já defendiam as coletas de dados para validação e análises dos mesmos. De outra forma, segundo Roosendaal e Geurts (1997), a pesquisa científica poderia ser uma vítima da chamada crise de informação, e tal crise poderia retardar o progresso da ciência como todo.

O ponto de passagem obrigatório (PPO) nessa perspectiva é a Disponibilidade de dados de Pesquisa, ou seja, condição vinculante de diversos atores para um objetivo dentro da rede. A proposta dos dados de pesquisa como PPO no âmbito da rede de comunicação científica, composta de atores, seus obstáculos e finalidades nas interações com os dados de pesquisa é ilustrada na Figura 6. É a ação em comum para os atores, ainda que esses possuam obstáculos distintos (letramento, conformidade, armazenamento de dados, entre outros). De acordo com as premissas de Grigg (2016), nas relações dos diversos atores com o PPO há que se considerar as

múltiplas interações de acordo com a tipologia das agências dos atores. O PPO possibilita maior visibilidade e o aumento das citações da pesquisa, atendimento às regras e exigências de financiadoras/revistas, conforme descrito na Figura 6.

Figura 6 - Disponibilidade dos dados de pesquisa como ponto de passagem obrigatório



Fonte: Autoria própria (2019) com base no modelo de Callon (1986) e nas premissas de Grigg (2016).

Do aspecto do Pesquisador na qualidade de leitor, os obstáculos passam pelo Letramento em dados - localização, reúso (proveniência, qualidade, extração, análises), visando o Reúso de dados, a pesquisa colaborativa e a aceleração da Ciência, conforme posicionado na Figura 6. As Agências de Fomento precisam da conformidade dos pesquisadores com suas políticas que objetivam a eficiência do investimento, a transparência da destinação dos recursos, a redução da duplicação, além de evitar fraudes e promover boas práticas de pesquisa.

As Revistas Científicas (revisores, editores) enfrentam as questões relativas ao armazenamento de dados, a curadoria de dados, a capacidade de revisão por pares, buscando o benefício da verificabilidade da Pesquisa, da revisão na íntegra dos artigos submetidos, aumentando a sua credibilidade. As universidades e Instituições de Pesquisa, ainda vinculadas a um modelo de recompensa baseado no volume de produção científica, submetida às avaliações de órgãos de controle, teriam nos dados de pesquisa em acesso aberto, a possibilidade de reprodução da pesquisa, da aceleração da Ciência, e do cumprimento das boas práticas de pesquisa. As Bibliotecas Acadêmicas seriam beneficiárias da disponibilidade dos dados, pois isso contribuiria para que mantivessem seu papel relevante junto às necessidades dos Pesquisadores. Dessa forma, estariam constantemente alinhadas às atividades-fim da instituição na qual estão inseridas, incrementando os serviços de apoio e suporte à sua comunidade, buscando manter seu quadro de profissionais habilitados para novos desafios. O Governo que precisaria prever Políticas Públicas e manejar múltiplos interesses para a disponibilidade de dados, contribuiria com a governança, a democratização da Ciência nos moldes da Ciência Aberta. Para Machado (2015), no plano político, os benefícios da disponibilidade de dados são mais profundos ainda, pois este reitera que o direito de acesso à informação é considerado um direito humano fundamental e que as iniciativas de abertura na ciência têm contribuído com a construção de novos paradigmas na distribuição do conhecimento científico. A Sociedade que se beneficiaria com a perspectiva da Ciência Cidadã, teria maior acesso aos resultados de interesse social, mas encontraria no Letramento em dados um obstáculo, ainda que esses estejam disponíveis.

A TAR não é, portanto, uma configuração fechada (SAYES, 2014). Segundo Sayes (2014), a controvérsia em torno da agência atribuída aos não-humanos está no fato da compreensão parcial do real significado desta agência. A partir das concepções que o termo não humano representa na TAR, remete-se a uma crítica à tradição filosófica da relação sujeito-objeto sendo tratada de forma radicalmente distinta ou polarizada. Termos como ator, actante, mônada, quasi-objeto ou ator híbrido, que apesar de se referirem a mesma insatisfação com a posição da filosofia tradicional, são termos distintos que estão sob o guarda-chuva de não-

humanos e capturam a essência das entidades que existem, promovendo a dissolução do natural e do social (SAYES, 2014).

O trabalho da tecnociência está em compreender os interesses de variados atores, traduzi-los para que todos trabalhem em conjunto e em acordo (SISMONDO, 2010), de forma simétrica, em um processo denominado sociologia da tradução (CALLON, 1986), conforme já citado. A noção de tradução compreende o processo de contínua mudança de objetivos e interesses que ocorrem nos diversos estágios das relações entre natureza e sociedade (CALLON, 1986).

Nesta perspectiva, na próxima seção discute-se as funções da Comunicação Científica e a rede composta por diversos atores que exercem influência em um processo colaborativo, que extrapola a comunicação dos resultados das atividades científicas.

2.5 FUNÇÕES DA COMUNICAÇÃO CIENTÍFICA

Para Meadows (1998), a razão fundamental da pesquisa é ser comunicada aos outros. A divulgação dos resultados em si, portanto, não é a única finalidade do processo de comunicação científica. Roosendaal e Geurts (1997) categorizaram quatro funções essenciais da comunicação científica: (i) o registro, em que o autor assegura que ele é a pessoa responsável por aquela parte da pesquisa, a pessoa que fez uma descoberta específica; (ii) a certificação, que passa pelo processo de revisão por pares e determina se o que o autor argumenta é razoável (o que também favorece os leitores); (iii) a divulgação, na qual a pesquisa é comunicada; e (iv) o arquivamento, em que a pesquisa é registrada para a posteridade.

Borgman (2007, p. 66) agrupa essas funções em (a) legitimidade; (b) disseminação; (c) acesso, preservação e curadoria. Guedon (1994, p. 4) também considera na sua classificação a “legitimidade e autoridade”. A ideia de legitimidade está ligada à noção de “confiabilidade” de Kling (2004, p. 594), pois considera o controle de qualidade do documento em relação ao conteúdo, ao atendimento de normas ou a reputação de um autor ou de sua instituição (KLING, 2004, p. 594). A disseminação engloba a categoria “comunicação e difusão” (GUEDON, 1994, p. 4), a “divulgação” (ROOSENDAAL; GEURTS, 1997, p. 17), ou a “Publicidade” (KLING, 2004, p. 594).

Os elementos “acesso, preservação e curadoria” (BORGMAN, 2007, p. 67) estão relacionados ao conceito de “arquivamento” (ROOSENDAAL; GEURTS, 1997, p. 17), de “memória” (GUEDON, 1994, p. 4) e de “acessibilidade” (KLING, 2004, p. 594). Acesso implica permissão e direito de uso, conexão às redes de computadores e também habilidades

para reuso de algo adquirido (BORGMAN, 2007, p. 67). Curadoria implica atividades de preservação e arquivo em ambientes digitais (BORGMAN, 2007, p. 67).

Os conceitos utilizados pelos autores citados para a descrição das funções da comunicação científica são comparados no Quadro 2.

Quadro 2 - Comparativo dos conceitos das funções da comunicação científica

Borgman (2007)	Kling (2004)	Roosendaal e Geurts (1997)	Guedon (1994)
Legitimidade	Confiabilidade	Registro	Legitimidade
		Certificação	Autoridade
Disseminação	Publicidade	Divulgação	Comunicação e difusão
Acesso, preservação e curadoria		Arquivamento	Memória
	Acessibilidade	-	-

Fonte: Adaptado de Borgman (2007), Guedon (1994), Kling (2004) e Roosendaal e Geurts (1997).

A comunicação da produção científica, portanto, está centrada na divulgação dos resultados das pesquisas, atrelados às questões de autoria, de proveniência e de preservação da informação. Contudo, não basta que a publicação de um determinado conhecimento esteja disponível, este precisa ser absorvido (MEES, 1918). Mees (1918) defendeu essa ideia quando escrevia para a revista *Nature* há quase um século. Esse autor identificou três estágios da produção do conhecimento científico: primeiro a produção de novos conhecimentos por meio de laboratório de pesquisa; em segundo lugar a publicação deste conhecimento na forma de trabalhos e resumos de trabalhos; em terceiro lugar a “digestão” do novo conhecimento e sua absorção de informações por comparação crítica com outras experiências sobre assuntos idênticos ou similares (MEES, 1918, p. 355).

Corroborando Mees (1918) e Meadows (1998), Latour (2000) também descreve a comunicação científica como o meio de divulgação e de publicidade dos fatos científicos, sendo que a forma dessa comunicação é o texto científico (LATOUR, 2000). Logo, a comunicação científica depende fortemente da divulgação de artigos, de artigos de revistas e monografias, inicialmente em formato impresso (PEPE *et al.*, 2009). Porém, no início dos anos 1990, o surgimento da internet possibilitou outras opções de comunicação e publicação, começando com a distribuição de *e-mail*, depois mudando para os outros tipos de comunicação *online* (CULLEN; CHAWNER, 2011).

Atualmente, os artefatos científicos originam-se já em forma digital e existem em uma variedade muito maior de formatos. Estes incluem manuscritos, publicações, dados, laboratório e notas de campo, calibrações de instrumentos, pré-impressões, propostas de concessões, palestras, *slides*, pedidos de patentes, teses, dissertações e tipos específicos para determinadas

disciplinas, e nem cientistas nem bibliotecários estão lidando bem com este “dilúvio de dados” como indicado por Pepe *et al.* (2009, p. 567), que tem afetado todas as disciplinas (WOOD, 2013).

O termo “dilúvio de dados” foi cunhado por Hey e Trefethen (2003, p. 812) que procuraram sintetizar a ideia de um crescimento exponencial na geração de dados científicos. Pepe *et al.*, (2009, p. 567), assim como Mees (1918), questionam como “digerir e absorver o dilúvio digital de artefatos na massa geral de informações de hoje”, visto que a necessidade de comparação crítica de Mess (1918) ainda permanece essencial.

Ainda que haja ferramentas sofisticadas que facilitem a comparação e a análise, uma gama mais ampla de artefatos acadêmicos está publicamente disponível e esses artefatos geralmente existem em múltiplas versões ou múltiplos estágios de desenvolvimento (PEPE *et al.*, 2009). A escola democrática da Ciência Aberta parte da premissa que o acesso ao conhecimento não é igualmente distribuído, logo, o acesso aberto e os dados abertos são ferramentas que podem contribuir para que o conhecimento seja disponível gratuitamente a qualquer um. As bases desta escola são os alicerces para a disponibilidade de dados abertos (FECHER; FRIESIKE, 2014).

2.6 ACESSO ABERTO – GÊNESE, DESENVOLVIMENTO E BENEFÍCIOS

A trajetória do acesso aberto no mundo passa por marcos importantes como: a Open Archives Initiative, em 1999 na Santa Fé Convention; a declaração de independência pela Scholarly Publishing & Academic Resources Coalition (SPARC) em 2001; a criação da Public Library of Science (PLOS), em 2001, a Declaração de Budapeste em 2002 no evento do Open Society Institute, as declarações de Berlin, Bethesda e Unesco em 2003; a Declaração da International Federation of Libraries Association (IFLA) sobre livre acesso a informação científica e tecnológica, em 2003; a Declaração da Organization for Economic Cooperation and Development (OECD) sobre o acesso aos dados de pesquisa financiada por fundos públicos, em 2004, o manifesto em prol do acesso aberto no Brasil, em 2005. Todos esses grupos têm advogado pela causa do acesso aberto, como afirmam em seus respectivos manifestos, e vêm fortalecendo o movimento em nível transacional, já que representam diversos países.

Porém, a primeira política de uma instituição de pesquisa formalizando o acesso aberto só veio em 2005 (KILEY, 2013). Diante da cobrança de uma taxa de 30 dólares para acesso a um artigo sobre a vacina da malária, pesquisa a qual tinham financiado em 2002, o diretor da Welcome Trust se deu conta de que nem ele nem as demais pessoas interessadas (da área médica

ou não) poderiam ter acesso aos resultados da pesquisa sem o pagamento da assinatura ou do artigo. Três anos depois, em 2005, a Wellcome Trust foi a primeira agência de financiamento na Europa a lançar sua política de acesso aberto (KILEY, 2013).

Segundo o Open Aire, projeto que tem como missão apoiar as políticas de acesso aberto e de dados abertos na Europa e no mundo, acesso aberto é a “disponibilidade imediata, *online* e gratuita de resultados de pesquisa sem restrições de uso normalmente impostas pelos acordos de direitos autorais do editor” (OPEN AIRE, 2015, *online*, tradução nossa). O acesso aberto inclui os resultados que os cientistas normalmente oferecem gratuitamente para a publicação; incluem artigos de revistas revisados por pares, documentos de conferência e conjuntos de dados de vários tipos.

Outras agências de financiamento no Reino Unido também estabeleceram políticas e princípios sobre o acesso aberto e a publicidade de pesquisa financiada com recursos públicos, tais como: UK Research Council (2005), Medical Research Council (2006), Biotechnology And Biological Science Research Council (2008) (KILEY, 2013). Ainda na Europa, inúmeras agências seguiram essa mesma abordagem, tendo como marco o projeto piloto Open Aire, em 2008 (OPEN AIRE, 2015). A partir de 2014 instituiu-se que todos os resultados de projetos europeus financiados no âmbito do projeto Horizon 2020 (EUROPEAN COMMISSION, 2014) estariam disponíveis em acesso aberto – inclusive os dados de pesquisa.

Nos EUA o movimento pelo acesso aberto moldou as políticas de agência de financiamento a partir de 2004 com a National Institutes of Health (NIH) (2003), ainda como um requisito e não como mandato, sem qualquer sanção para quem não seguisse a recomendação. Após uma legislação aprovada no congresso norte-americano em 2007, transformando a recomendação em exigência, a taxa de adesão passou de 4% para 75%. Essa normativa incentivou outras agências federais a regularem o acesso aberto como condição para o investimento em pesquisa nos Estados Unidos (KILEY, 2013).

Como benefícios do acesso aberto, estão inclusos: (a) visibilidade: foi demonstrado por Davis *et al.* (2008) que os artigos em acesso aberto tiveram 89% mais acesso ao texto completo que aqueles em acesso restritos; (b) melhora no processo de pesquisa: a publicação aberta permite o acesso ao texto para mineração de dados e aplicativos de extração de informações; (c) redução de custos: mesmo com o modelo de pagamento pela via dourada (modalidade em que custos com a publicação são pagos pelo autor ou pelo instituto de financiamento), gera-se uma economia em escala em relação aos custos de acesso por assinaturas de periódicos. Em valores de 2010, a Wellcome Trust investia em torno de 1% de seu orçamento em taxas de depósitos para publicações em acesso aberto, disponibilizando os resultados de forma ampla e

garantindo o acesso de cada indivíduo aos resultados científicos (KILEY, 2013).

Manter os elevados custos de assinaturas também têm sido um desafio para as bibliotecas. Bosch e Henderson (2012) mostram como os orçamentos das bibliotecas de universidades americanas têm sido reduzidos ano após ano, enquanto que os preços médios dos títulos de periódicos e até de artigos individuais tem aumentado constantemente dentro do modelo de pacote de dados que são oferecidos pelas editoras, que constituem o oligopólio das publicações científicas. Um panorama histórico dessa discrepância foi apresentado pela Association of Research Libraries (ARL), mostrando que os valores de assinaturas de periódicos científicos cresceram em 402% entre os anos de 1986 a 2011 (ASSOCIATION OF RESEARCH LIBRARIES, 2012).

Em oposição ao acesso restrito está o acesso aberto e para viabilizar o acesso aberto à produção científica, duas mudanças contribuíram com a democratização do acesso aos textos produzidos nas pesquisas. A primeira diz respeito ao movimento pelo *software* livre de código fonte aberto, que viabilizou a infraestrutura tecnológica necessária ao livre acesso. Iniciado ainda na década de 1970, após diversos desdobramentos do movimento, o *software* livre foi o que favoreceu o desenvolvimento das ferramentas para publicação e disponibilização da produção científica de forma aberta e gratuita (EUROPEAN COMMISSION, 2006).

A segunda, mais recentemente, refere-se às mudanças políticas das agências de fomento que têm promovido um ambiente social favorável à democratização da ciência, na medida em que reduz ou anula os custos de acesso à pesquisa financiada com recursos públicos e/ou privados. Nesse sentido os órgãos de financiamento têm exercido um papel central na apropriação da ideia do acesso aberto, sendo catalisadores da mudança, mostrando que o modelo de assinatura dos periódicos não tem mais atuado pelos melhores interesses da comunidade acadêmica ou mesmo da sociedade (KILEY, 2013).

Segundo o Open Science and Research Initiative (2014), a principal premissa da pesquisa aberta é a democratização da ciência. O acesso aberto aos artigos acadêmicos impulsionou as pesquisas na área de comunicação científica nos últimos anos, considerando que esta forma de acesso contribuiu para o aumento da disponibilidade de conteúdos em formatos eletrônicos (CREASER, 2011). O acesso aberto tem sido discutido em diversos países e o Brasil detém um dos maiores percentuais de publicações científicas em acesso aberto do mundo (SCIENCE-METRIX, 2018).

2.7 ACESSO ABERTO NO BRASIL

Uma pesquisa do Science-Metrix (2018), identificou que 75% dos artigos em periódicos nacionais brasileiros estão disponíveis gratuitamente, a maior parte hospedada na biblioteca *online* Scientific Electronic Library Online (SCIELO). Isso representa que o Brasil tem o maior percentual de acesso aberto dentre os países pesquisados, seguido pela Holanda, também com 75% de percentual de artigos em acesso aberto e a Suíça, em terceiro lugar, com 70% (SCIENCE-METRIX, 2018). Os marcos brasileiros do acesso aberto são o Manifesto pelo Livre Acesso do Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict) em 2005 (INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA, 2005), assim como a Declaração de Salvador durante o Ninth World Congress on Health Information and Libraries, Salvador, no mesmo ano (KURAMOTO, 2006, SAYÃO; MARCONDES, 2009). O Manifesto brasileiro apresenta “um conjunto de recomendações a todos os segmentos da comunidade científica para a implantação do acesso livre no País” (KURAMOTO, 2009, p. 212).

Em setembro de 2016 o Ibict estendeu seu Manifesto apresentado em 2005 em que advoga pelo acesso aos dados abertos de pesquisa, sob o título Manifesto de Acesso Aberto aos Dados de Pesquisa Brasileira para Ciência Cidadã (INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA, 2016, *online*); neste documento o Ibict “estende a sua visão sobre o acesso aberto, e reconhece os dados de pesquisa como um recurso imprescindível para as ações de Ciência Aberta, Ciência para todos, Ciência Cidadã”. O manifesto recomenda diretrizes gerais às seguintes entidades:

[...] universidades e institutos de pesquisa brasileiros, às sociedades científicas e academias de ciência, aos órgãos de fomento à pesquisa, aos editores de revistas ou periódicos científicos, aos cursos de pós-graduação e graduação nas áreas de informação, aos gestores e executores de programas e projetos de dados de pesquisa e aos pesquisadores. (INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA, 2016, *online*).

Outra iniciativa coordenada pelo Ibict é a aprovação de uma política institucional nas instituições de ensino superior (IES), por força do edital Ibict-FINEP/PCAL/XBDB que regula o depósito em acesso aberto pelos pesquisadores (INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA – IBICT, 2009). Fatores como a disponibilidade de *software* de código fonte abertos, além de diversos recursos na internet contendo metodologias e casos de sucesso como exemplos, contribuíram para a facilidade de criação de repositórios institucionais (RIs), sendo vários deles implementados no Brasil por

conta desses incentivos (SETENARESKI; SUNYE; SHIMA, 2013). Com 152 RIs, o país é o líder com 38% dos 392 RIs da América Latina, segundo o Registry of Open Access Repositories – ROAR (2018), uma base pesquisável que indexa a criação, a localização e o crescimento de repositórios institucionais de acesso aberto e seus conteúdos em todo o mundo.

A partir do cenário legislativo nacional, infere-se que uma questão, porém, permanece desafiadora: a construção de políticas públicas para regular o acesso aberto às publicações que são resultados de pesquisas financiadas com recursos públicos no Brasil, sem barreiras financeiras e/ou tecnológicas. Apesar de já terem sido apresentados no Congresso Nacional projetos de lei concernentes à divulgação da produção acadêmica em acesso aberto por meio de RIs, ainda não existe no Brasil uma legislação abrangente que regule o acesso aberto aos resultados de pesquisa das IES públicas, como pode ser visto a seguir nos exemplos de propostas legislativas que não foram aprovadas.

2.7.1 Projetos de Lei em Nível Nacional sobre Acesso Aberto no Brasil

Prevendo a necessidade de uma legislação para a regulação da publicação científica brasileira, o Projeto de Lei PL 1120/2007 (BRASIL, 2007, *online*) estabelecia a obrigação das instituições públicas de ensino superior construir os repositórios institucionais para depósito da produção acadêmica do corpo discente e docente. Este projeto tramitou no congresso Nacional brasileiro de 2007 até 2012 (BRASIL, 2012a, *online*), quando foi arquivado sem aprovação porque o Deputado autor do projeto, Rodrigo Sobral Rollemberg, não foi reeleito, como exige o regimento interno (BRASIL, 2012b, *online*).

Uma nova proposta em favor do acesso aberto de Rodrigo Sobral Rollemberg, então como Senador, é o PLS 387/2011 (BRASIL, 2011, *online*), basicamente nos mesmo moldes do projeto apresentado enquanto era Deputado. Entretanto, é o pesquisador Hélio Kuramoto que tem advogado pela causa de uma legislação para acesso aberto junto aos políticos brasileiros. Atualmente Professor na Universidade Federal de Minas Gerais – UFMG, o pesquisador é graduado em Engenharia e Doutor em Ciências da Informação e da Comunicação. Ele tem advogado pela causa do acesso livre à informação científica de forma mais expressiva nos últimos 10 anos. Em sua entrevista para Portela (2013, *online*), ele destaca: “Em seguida, articulamos com o atual senador da República Rodrigo Rollemberg, o mesmo que submeteu o PL 1120/2007, a submissão de um projeto similar no Senado”.

Então, desde 2011, o PLS 387/2011 (BRASIL, 2011, *online*) estava em discussão. Em

consulta realizada em julho de 2019, a última posição informada é de que a proposta foi arquivada ao final da Legislatura, em 27/12/2018.

Kuramoto (2013) ainda deixa claro que as articulações junto ao Senador, tem sido resultado de um esforço praticamente individual, sem o envolvimento do Ibict, do qual era pesquisador. “O PLS 387/2011 não foi submetido pelo Ibict. O nome do Ibict aparece, apenas, pelo fato de à época eu ser funcionário daquele Instituto, mas, a sua direção nada tem feito em favor do referido projeto de lei.” (KURAMOTO, 2013, *online*).

Outra tentativa de regulamentar os RIs na Câmara dos Deputados se deu pelo Projeto de Lei PL 6702/2013 (BRASIL, 2013b, *online*), que está em tramitação desde 2013 (BRASIL, 2013a, *online*). Apesar do primeiro projeto de lei (PL 1120/2007) ter sido aprovado e recomendado por várias comissões da Câmara dos Deputados (BRASIL, 2009, *online*), o segundo projeto (PL 6702/2013) que trata do mesmo teor do primeiro, teve seu parecer inicial rejeitado pela relatora Deputada Margarida Salomão (PT-MG) (BRASIL, 2014a, *online*).

Uma das justificativas para a rejeição é que o Ibict já mantém a Biblioteca Digital Brasileira de Teses e Dissertações – BDTD, consolidando nacionalmente essas informações. Porém, Kuramoto (2009) defende a necessidade de ser ter uma legislação ou política que torne obrigatório o registro da produção científica das instituições, o que atualmente não ocorre. Não há qualquer previsão de que as instituições atualizem os seus repositórios, ou seja, a BDTD também não tem a garantia da inserção da totalidade da produção de teses e dissertações em seu repositório. Segundo Kuramoto (2009), outro motivo para dificuldade do estabelecimento de qualquer mandato ou obrigatoriedade no processo de registro e disseminação da produção de teses e dissertações no país é que não existe qualquer autoridade do Ibict sobre as instituições de ensino superior. Se a disseminação das teses e dissertações não tem uma previsão legal, da mesma forma a publicidade e a disseminação dos dados de pesquisa não está sujeita a uma legislação nacional. É exatamente a autonomia universitária, citada por Kuramoto (2014), que motivou o Senador Álvaro Dias, como relator, a votar pela rejeição do projeto PLS 387/2011 em 2013 (BRASIL, 2013c, *online*).

Outro argumento é de que “bases de dados abrangentes de apoio à pesquisa e ao desenvolvimento tecnológico são oferecidas pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES, no seu portal de periódicos[...]” (BRASIL, 2014, p. 2). Entretanto, 50% das 520 bases de dados do conteúdo do Portal de Periódicos da Capes é referente às assinaturas de periódicos científicos publicados por editores comerciais.

Esta objeção só demonstra o desconhecimento da relatora desse projeto sobre o cenário e os contextos brasileiros envolvendo a divulgação da produção acadêmico-científica das IES.

Portanto, nota-se que Kuramoto (2014) tem exercido pouca influência diretamente nas propostas de lei de acesso aberto no Brasil. Uma de suas ações foi a articulação com a classe política brasileira com o objetivo de aprovar no Senado Federal o PLS 387/2011, projeto de lei submetido pelo Senador Rodrigo Rollemberg, a quem apoiou na elaboração (KURAMOTO, 2014). Ainda que não tenha sido aprovada uma norma nacional que legisle sobre o acesso aberto às publicações e/ou aos dados de pesquisa, as instituições de ensino e de financiamentos elaboram suas próprias normas internas para orientar as suas comunidades acadêmicas quanto ao nível e aos critérios de exigências das pesquisas que financiam. Algumas dessas políticas começam a ser elaboradas nas instituições de pesquisas do Brasil, contudo.

2.7.2 Políticas em Níveis Institucionais de Acesso Aberto no Brasil

Muitas políticas institucionais em acesso aberto têm sido elaboradas por instituições de ensino e agências financiadoras como forma de atender as exigências de leis sobre o tema em vários países (BRINEY; GOBEN; ZILINSKI, 2015; CENDI, 2016). De acordo com o The Registry of Open Access Repository Mandates and Policies – ROARMAP (2018), que faz o registro internacional sobre regulamentos internos e políticas adotadas para o acesso aberto por IES por meio dos seus repositórios institucionais, existem 898 políticas mandatórias, de diversos tipos e finalidades aprovados pelos países. Na América do Sul estão 45 (5%) destes regulamentos, sendo que 22 (47%) do bloco pertencem ao Brasil (ROARMAP, 2018).

O ROARMAP (2018) faz uma análise dos critérios de cada regulamento em relação ao padrão do Programa Europeu H2020, atribuindo um percentual de alinhamento com o mesmo. O mapa, à esquerda da Figura 7, apresenta o percentual de aderência das políticas de acesso aberto dos países representados em relação à política do programa H2020. Em média, as políticas sobre acesso aberto no Brasil estão alinhadas em 52% com o padrão H2020, como pode ser visto na Figura 7 (ROARMAP, 2018). A Figura 7 também apresenta, ao lado direito, um gráfico com os percentuais de aderência de 196 políticas institucionais, separadas por tipo de legislador, de acordo com a seguinte legenda: Azul – Financiador; Laranja – Financiador e Organização de Pesquisa; Vermelho - Organização de Pesquisa e Lilás – Subunidade de Organização de Pesquisa.

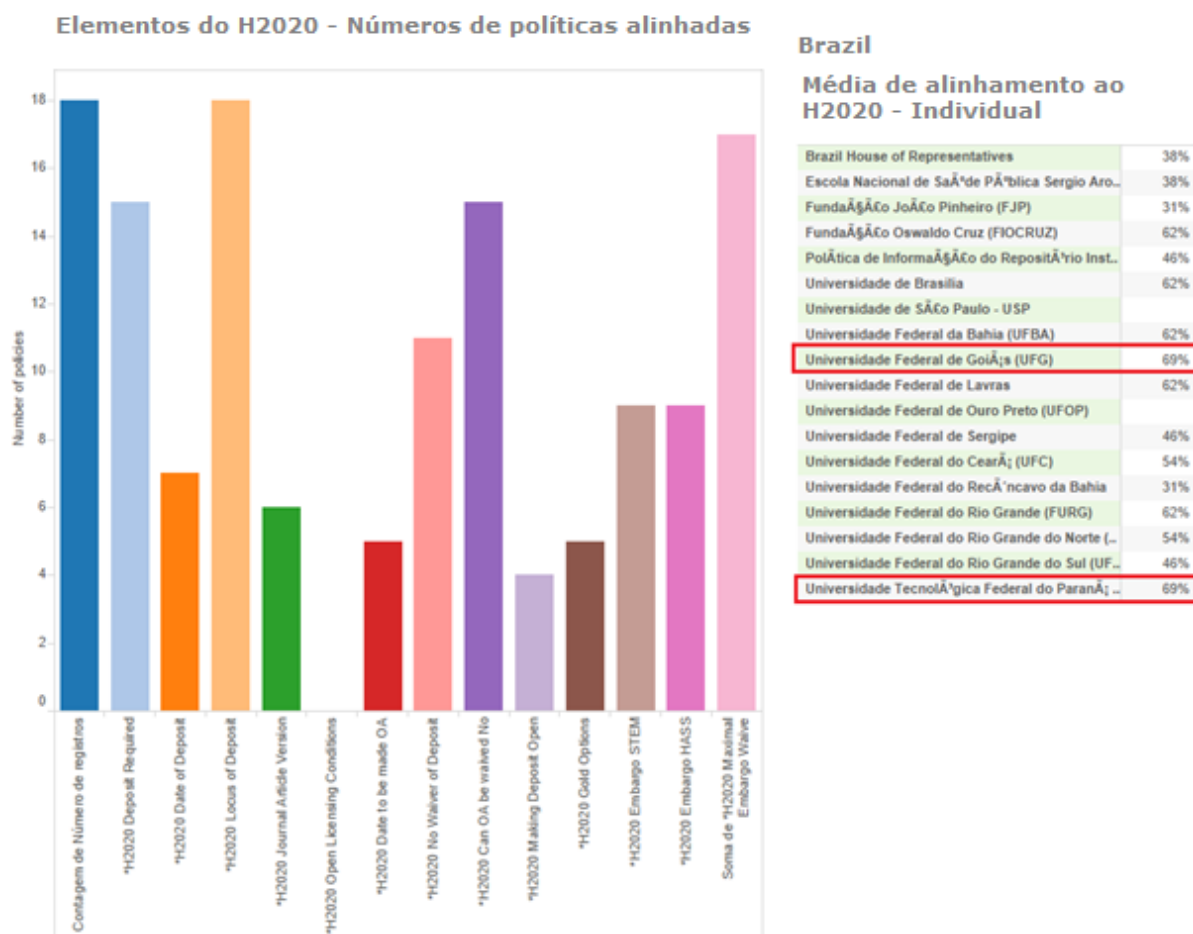
Figura 7 - Percentual de alinhamento dos regulamentos em acesso aberto brasileiros ao H2020



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

A Universidade Tecnológica Federal do Paraná, junto com a Universidade Federal de Goiás são as duas instituições federais brasileiras que possuem o maior nível de aderência ao padrão de excelência proposto pelo ROARMAP (2018): 69% de alinhamento com o H2020, como mostra a Figura 8, no detalhamento do nível de aderência das políticas individualmente, com destaque para as duas universidades nacionais.

Figura 8 - Nível de aderência de regulamento individual ao padrão de excelência H2020



Fonte: ROARMAP (2018, *online*).

As revistas acadêmicas brasileiras também têm elaborado as próprias políticas para o acesso aberto. Sendo assim, algumas políticas de revistas nacionais e quais as permissões para seus conteúdos estarem disponíveis em acesso aberto, são apresentadas a seguir, contrapondo a disponibilidade e o acesso das produções acadêmicas nos RIs.

2.7.3 Políticas de Acesso Aberto das Revistas Científicas Brasileiras

Os periódicos acadêmicos têm políticas emergentes sobre compartilhamento e acesso a dados que também afetam os pesquisadores em todo o mundo (BRINEY; GOBEN; ZILINSKI, 2015; NATURE PUBLISHING GROUP, 2014; PLOS ONE, [s.d.]; PUBLIC LIBRARY OF SCIENCE, [s.d.]). Um panorama das políticas das revistas científicas brasileiras pode ser visualizado no Diretório de Políticas de Acesso Aberto das Revistas Científicas Brasileiras – DIADORIM (INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E

TECNOLOGIA, 2018, *online*). O serviço permite localizar informações sobre a política editorial das revistas científicas brasileiras em relação à permissão para o armazenamento e o acesso de artigos em acesso aberto. Esta é a versão brasileira dos serviços de políticas de editores como o SHERPA RoMEO para editoras de todo o mundo (SHERPA RoMEO, *online*), Dulcinea, que cobre editores da Espanha (DULCINEA, *online*); Blimunda para editores de Portugal (DOCUMENTATION AND LIBRARY DIVISION OF THE FACULTY OF SCIENCES AND TECHNOLOGY, 2018) e OAKLIST para editores da Austrália (OAKLIST, *online*).

Os dados apresentados no *site* Diadorim classificam as políticas de depósito e de acesso das 1.547 editoras cadastradas no Diretório. A classificação de depósito é organizada por cores: a classe branca é a mais restrita, pois apresenta restrições para o depósito das versões *preprint* e *pós-print* de um artigo. Já a classe amarela significa que a revista permite o depósito da versão *preprint* (versão do autor – manuscrito digital) de um artigo. A classe azul permite o depósito da versão *pós-print* (versão de publicação do artigo, avaliada pelos pares e corrigida pelos autores e o PDF da Revista) de um artigo. A classe verde permite o depósito de ambas às versões *preprint* e *pós-print* de um artigo. Em termos de depósito, 1.531 (97%) das revistas permitem que sejam armazenados algum tipo de versão de seus artigos, que seria a soma das classes amarela, azul e verde. Apenas 16 (1%), a classe branca, não permitem o depósito (INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA, 2018, *online*).

Ainda que as permissões de depósito e de acesso sejam concedidas de formas separadas, 1.494 (96%) das políticas não preveem qualquer tipo de impedimento ao acesso, seja por embargo ou restrição, portanto, são classificadas como acesso aberto (Tabela 3). O segundo nível apresentado na Tabela 3 é o acesso com embargo, uma imposição de um período de tempo para que o artigo seja acessado, e representa o tipo de acesso a 39 políticas (3%). Já no acesso restrito, nenhum usuário poderá acessar o documento depositado em qualquer tempo. Apenas 14 (1%) das revistas brasileiras restringem o acesso aberto (INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA, 2018, *online*).

Tabela 3 - Políticas de depósito e acesso das revistas científicas brasileiras

Depósito			Acesso		
Tipo	Qtde	%	Tipo	Qtde	%
Amarelo	41	3%	Aberto	1494	96%
Azul	903	58%	Com embargo	39	3%
Branco	16	1%	Restrito	14	1%
Verde	587	38%			
Total	1.547	100%		1.547	100%

Fonte: Adaptado de Instituto Brasileiro de Informação em Ciência e Tecnologia (2018).

Legenda:

Amarelo	Permite o depósito da versão <i>preprint</i> (versão do autor – manuscrito digital) de um artigo
Azul	Permite o depósito da versão <i>pós-print</i> (versão de publicação do artigo, avaliada pelos pares e corrigida pelos autores e o PDF da Revista) de um artigo
Branco	Apresenta restrições para o depósito das versões <i>preprint</i> e <i>pós-print</i> de um artigo
Verde	Permite o depósito das versões <i>preprint</i> e <i>pós-print</i> de um artigo

Esses dados demonstram que as revistas científicas brasileiras se posicionaram favoravelmente ao acesso aberto. O principal canal de acesso das revistas é a internet, por meio de plataformas *online* disponibilizadas, na sua maioria, nos *sites* e nos repositórios institucionais hospedados nas bibliotecas universitárias. As bibliotecas têm tido um papel fundamental neste processo nos últimos 150 anos e mais intensamente, na última década em relação ao acesso aberto, viabilizando repositórios institucionais que promovem a disseminação aberta do conhecimento científico (CULLEN; CHAWNER, 2011). Como responsáveis pela gestão da informação acadêmica produzida na instituição, seja por meio de coleções físicas ou, atualmente das coleções digitais, o bibliotecário, para Ortega Y Gasset (1967) torna-se um importante colaborador, tanto do cientista quanto do pesquisador. Tradicionalmente as bibliotecas têm atuado no papel de provimento do acesso e da disseminação da produção científica de várias formas, portanto, a questão dos dados de pesquisa tem sido uma pauta global para as bibliotecas universitárias (TENOPIR *et al.*, 2017; TENOPIR; BIRCH; ALLARD, 2012; TENOPIR *et al.*, 2015a). Creaser (2011) também destaca o papel das bibliotecas acadêmicas dentro do processo de comunicação científica como um provedor de informações e como facilitador do acesso.

Além do provimento ao acesso e funções de armazenagem, as bibliotecas acadêmicas também são importantes atores que têm desempenhado um papel na construção da cultura e da infraestrutura para os dados de pesquisa enquanto resultados de investigação científica

(KARDELIENE; KARDELIS; BAKUTYTE, 2014; TENOPIR *et al.*, 2015a).

Um conjunto de serviços de apoio aos pesquisadores para lidar com os dados de pesquisa tem se desenvolvido nas bibliotecas nos últimos anos e dentre esses serviços se destaca o suporte ao letramento informacional em dados (BROWN, 2013; KARDELIENE; KARDELIS; BAKUTYTE, 2014; KOLTAY, 2017b; TENOPIR; BIRCH; ALLARD, 2012; TENOPIR *et al.*, 2013; TENOPIR *et al.*, 2016; TENOPIR; HUGHES; ALLARD, 2015). Os dados de pesquisa e o letramento informacional são o objeto do próximo Capítulo.

3 OS DADOS DE PESQUISA E O LETRAMENTO INFORMACIONAL

Este capítulo apresenta a terminologia e a conceituação em torno dos dados de pesquisa e como o letramento em dados pode ajudar a mitigar os desafios decorrentes do volume exponencial de dados, por um lado, e da exigência cada vez maior da disponibilidade dos mesmos, por outro lado. Discute-se como o letramento informacional de dados tem sido principalmente viabilizado pelos serviços ofertados nas Bibliotecas Acadêmicas. Essas entidades têm apoiado a tarefa de suporte da pesquisa orientada a dados e têm adotado diversas abordagens na oferta desses serviços. Finaliza-se apresentando *frameworks* de letramento em dados a partir da literatura e da prática de entidades supra institucionais como Association of College and Research Libraries (ACRL), da Association of Research Libraries (ARL) e do Programa Facilitate Open Science Training For European Research (Foster). Algumas conceituações são elencadas de modo a ampliar a concepção dos dados de pesquisa.

3.1 O QUE SÃO OS DADOS DE PESQUISA

Para Wallis, Rolando e Borgman (2013), uma das maiores dificuldades em usufruir das potencialidades dos dados de pesquisa é justamente o entendimento da sua ontologia e natureza. Apesar de não haver um consenso e variar de acordo com as disciplinas, os dados de pesquisa, referem-se ao “material de fato registrado comumente aceito na comunidade científica como necessário para validar os resultados da pesquisa” (THE ENGINEERING AND PHYSICAL SCIENCES RESEARCH COUNCIL, 2018, *online*, tradução nossa). Cobrem uma ampla gama de tipos de registros e podem ser estruturados e armazenados em vários formatos de arquivos. “Embora a maioria desses dados seja criada em formato digital, todos os dados de pesquisa são incluídos na conceituação, independentemente do formato em que são criados” (THE ENGINEERING AND PHYSICAL SCIENCES RESEARCH COUNCIL, 2018, *online*, tradução nossa). Pesquisadores em quase todas as disciplinas criam dados em formato digital (HEY; TANSLEY; TOLLE, 2009).

Para a National Science Board (2005), os dados podem ser definidos como "qualquer registro que possa ser armazenado em formato digital, incluindo texto, números, imagens, vídeos ou filmes, áudio, *software*, algoritmos, equações, animações, modelos, simulações, etc." (NATIONAL SCIENCE BOARD, 2005, p. 9, tradução nossa). Esses dados podem ter muitas formas, por exemplo, as medidas registradas pelos satélites de monitoramento ambiental, os

produtos de colisões entre as partículas fundamentais, as sequências de genomas inteiros, os resultados das pesquisas das Ciências Sociais, as imagens de inscrições gregas antigas ou os vídeos anotados de danças (JISC SERVICES LIMITED, 2011). Em uma conceituação mais abrangente, são materiais gerados ou coletados durante a realização das pesquisas (NATIONAL ENDOWMENT FOR THE HUMANITIES, 2018).

Outras caracterizações abordam as finalidades dos dados de pesquisa: "Os dados de pesquisa, ao contrário de outros tipos de informações, são coletados, observados ou criados, para fins de análise para produzir resultados de pesquisa originais" (UNIVERSITY OF EDINBURGH, 2016, *online*, tradução nossa). São uma "representação reinterpretável da informação de forma formalizada, adequada à comunicação, interpretação ou processamento" (DIGITAL CURATION CENTER, 2007, *online*, tradução nossa). Ao caracterizar os dados, a Royal Society conceitua-os como "números, caracteres ou imagens que designam um atributo a um determinado fenômeno" (ROYAL SOCIETY, 2012, p. 12, tradução nossa). Da perspectiva ontológica, os dados são "descritivos do objeto de pesquisa, ou são o próprio objeto" (UNIVERSITY OF BATH, 2011, *online*).

Para a perspectiva deste trabalho, adota-se a conceituação do The Engineering and Physical Sciences Research Council (2018, *online*), pois esta não vincula a caracterização dos dados de pesquisa aos formatos de arquivos eletrônicos ou às formas de registro, mas sim à sua finalidade, que é a validação dos resultados da pesquisa, desde que aceitos dentro de uma determinada comunidade – o que demanda a validação dos pares. Salienta-se que os formatos e os processos de registro dos dados variam de disciplina para disciplina (WALLIS; ROLANDO; BORGMAN, 2013).

A National Science Foundation classifica os dados a partir de processos que lhe deram origem: dados observacionais, computacionais ou experimentais (NATIONAL SCIENCE BOARD, 2005). Os dados observacionais são gerados a partir de observações diretas, como medições meteorológicas e pesquisas de atitude, qualquer uma das quais pode estar associada a lugares e horários específicos ou envolvem múltiplos lugares e tempos (por exemplo, estudos transversais ou longitudinais) (NATIONAL SCIENCE BOARD, 2005). Os dados computacionais têm sua origem na execução de um modelo ou simulação de computador, sendo necessária a inclusão de uma descrição completa do *hardware*, do *software* e dos dados de entrada para permitir que o modelo seja replicado no futuro (NATIONAL SCIENCE BOARD, 2005). Os dados experimentais incluem resultados de estudos laboratoriais, como medições de reações químicas ou de experimentos de campo, tais como estudos comportamentais controlados (NATIONAL SCIENCE BOARD, 2005). O custo da reprodução desses

experimentos é um fator determinante para a sua preservação, sendo necessária a manutenção de documentação e de informações suficientes para reproduzir o experimento, caso os dados não sejam armazenados indefinidamente. Aqueles que não possam ser reproduzidos nas mesmas condições experimentais ou tiverem alguma variável desconhecida devem ser preservados no longo prazo (NATIONAL SCIENCE BOARD, 2005).

Wood (2013) reforça a importância do contexto aliado aos dados e seus significados. Sem a contextualização, apesar de disponíveis, os dados se tornam inúteis para o reuso. Nesse sentido considera-se o papel dos metadados para assegurar que os dados possam ser identificados e reutilizados. O relatório da Royal Society (2012) destaca que os dados devem não apenas estar disponíveis, mas precisam ser reutilizáveis por diferentes pessoas e em diferentes disciplinas. Portanto, o relatório destaca a importância de fazer com que os dados estejam não apenas acessíveis, mas também inteligíveis e utilizáveis para uma ampla gama de propósitos (ROYAL SOCIETY, 2012). Para a Royal Society (2012, p. 14, tradução nossa), a disponibilidade de dados que permite o reuso é conceituada como “disponibilidade inteligente”.

Considerando o contexto das Ciências Sociais, questões de privacidade de dados ampliam o desafio para o compartilhamento e o posterior reuso, visto que são dados mais centrados em participações humanas, aumentando os cuidados com a confidencialidade (EYNDEN; SCHRECKENBERG; CORTI, 2015, *online*). Assim como em outras disciplinas, os dados de pesquisa nas Ciências Sociais abrangem uma vasta gama de materiais, como por exemplo conjuntos de dados numéricos, entrevistas, notas de campo e documentos coletados para estudos etnográficos, podendo ser tanto quantitativos quanto qualitativos (CONSORTIUM OF EUROPEAN SOCIAL SCIENCE DATA ARCHIVES, 2018, *online*). Neste campo, provavelmente esses dados são provenientes de participantes humanos e isso pode significar acesso aos dados pessoais. Esses dados são considerados sensíveis e demandam cuidados especiais, principalmente porque uma pessoa específica pode ser identificada de forma direta ou indireta. Elementos que facilitam a identidade podem ser referências a um identificador como nome, número de identificação, dados de localização, ou a um ou mais fatores específicos da identidade física, fisiológica, genética, mental, econômica, cultural ou social de uma pessoa natural (CONSORTIUM OF EUROPEAN SOCIAL SCIENCE DATA ARCHIVES, 2018, *online*).

Seja na pesquisa quantitativa, em que as informações numéricas são coletadas e expressas de forma a serem avaliadas usando análises estatísticas, ou na pesquisa qualitativa, que reúne informações que não estão em formato numérico e são expressos em linguagem natural e posteriormente de forma visual, são necessárias medidas para proteger dados pessoais

(sensíveis) de acesso não autorizado; além da combinação entre a obtenção de consentimento, anonimização dados, mantendo a clareza sobre quem detém os direitos autorais de seus dados, é preciso controlar o acesso e o compartilhamento ético e legal desses dados pessoais (CONSORTIUM OF EUROPEAN SOCIAL SCIENCE DATA ARCHIVES, 2018, *online*).

Embora grandes quantidades de dados sejam vistas como uma solução para muitas questões de credibilidade de pesquisa são também consideradas um risco para a privacidade, pois diminuem a liberdade civil e inauguram um estado de aumento de controle corporativo (KOLTAY, 2015a). Nessa perspectiva, Koltay (2015a) atribui ao letramento em dados um papel de estímulo ao pensamento crítico que traga equilíbrio diante das possibilidades de uso dos dados de pesquisa, ou seja, evitar o excesso de otimismo pelo aumento do volume de dados e, ao mesmo tempo, considerar os desafios de privacidade e de confiabilidade decorrentes da disponibilidade dos dados (KOLTAY, 2015a).

3.2 O PROBLEMA DO EXCESSO DE DADOS E POSSÍVEIS SOLUÇÕES

A visão em expansão do que se entende por dados diante da crescente importância dos dados de pesquisa trouxe consigo o crescimento da necessidade de um letramento em dados. (KOLTAY; SPIRANEC; KARVALICS, 2016). Um exemplo é que os dados não são mais limitados aos resultados quantitativos provenientes de experimentos ou pesquisas (dados computacionais), desse modo os cientistas de Ciências Humanas e Sociais representam um novo público alvo para o letramento em dados, visto que produzem majoritariamente dados qualitativos (ASSOCIATION OF COLLEGE AND RESEARCH LIBRARIES, 2013).

Apesar de não ser um conceito novo e nem homogêneo, novos papéis têm influenciado a natureza do letramento em dados de forma extensiva (SCHNEIDER, 2013). Para Koltay (2015a), encontrar a identidade do letramento em dados requer o exercício de uma avaliação crítica saudável diante de uma crença generalizada de que a existência e o acesso aos dados de pesquisa, em particular, aos grandes volumes de dados, oferecem uma forma mais elevada de inteligência e de conhecimento (KOLTAY, 2015a).

Essa percepção crítica sobre o volume de produção e de publicação dos dados de pesquisa pode ser também identificada em autores que demonstram o lado negativo desse fenômeno, alertando sobre sintomas problemáticos como ansiedades, paradoxos e patologias de informação causadas pelo excesso desta disponibilidade. Hey e Trefethen (2003, p. 812) nomeiam esse problema como o “dilúvio de dados”, Bawden e Robinson (2009, p.180) chamam

de “sobrecarga de informação” e Awre *et al.*, (2015 p. 357) classificam este como um “problema perverso”.

Bawden e Robinson (2009) reconheceram que informações excessivas prejudicam o desempenho do pesquisador, pois a sobrecarga de informação gera um impedimento ao uso eficiente da informação. Essa sobrecarga obstrui a capacidade do pesquisador na filtragem de informações, dado o fosso cada vez maior entre o recurso informacional que está disponível e aquilo que é utilizável (KOLTAY, 2017c). A prevalência de dados de pesquisa nas Ciências Naturais, nas Ciências Sociais e nas Humanidades também contribui significativamente para a sobrecarga de recursos. Tomando a semelhança entre dados e informações, o volume excessivo de dados também pode ser chamado de sobrecarga de dados (KOLTAY, 2017c). Bawden e Robinson (2009) consideram que a sobrecarga emerge em função de um único fator, mas se origina por uma combinação de várias causas.

Essa ideia se aproxima da concepção de problema perverso que Awre *et al.*, (2015) emprestaram de Rittel e Webber (1973²³ *apud* AWRE *et al.*, 2015) para caracterizar os dados de pesquisa. Em uma conceituação breve, os problemas perversos ou indomáveis são aqueles distintos dos problemas comuns do dia-a-dia, aqueles que são rotineiros e a forma de intervenção é conhecida. Problemas perversos são muito mais complexos, pois são percebidos de forma muito distinta pelas partes interessadas envolvidas; não têm respostas fáceis e é difícil perceber o estágio em que são solucionados (AWRE *et al.*, 2015). Aplicado aos dados de pesquisa, estes foram caracterizados como um problema perverso devido à grande escala e complexidade das atividades de dados e de sua gestão, o amplo número de partes interessadas, a heterogeneidade dos tipos de dados e a falta de clareza nas funções apropriadas para as partes interessadas, incluindo as bibliotecas, bem como os serviços de suporte a serem oferecidos à comunidade acadêmica (TENOPIR *et al.*, 2017).

Para mitigar esses problemas de excesso de disponibilidade dos dados e a capacidade de manejo dos mesmos é que se insere o letramento em dados. Nesse sentido, as abordagens sociais estão disponíveis em uma escala muito mais ampla e variada do que as tecnológicas (KOLTAY, 2017c). Para Koltay (2017c), apesar de ter seus antecedentes e estar intimamente ligado ao letramento informacional, o letramento em dados traz uma nova faceta para o mundo das alfabetizações. Os quadros pragmáticos ou *frameworks* para essas atividades são oferecidos pelo letramento informacional e outros letramentos relacionadas que ajudam a entender melhor

²³ RITTEL, Horst. W; WEBBER, Melvin. M. Dilemmas in a general theory of planning. **Policy Sciences**, v. 4, n. 2, p. 155-169. 1973. Disponível em: <https://goo.gl/uZSSqL>.

o mundo digital, porque o que é digital também está sujeito à organização e à compreensão humana (ASSOCIATION OF COLLEGE AND RESEARCH LIBRARIES, 2000).

Os letramentos denominados informacional e de dados são socialmente mediados e as ferramentas tecnológicas moldam as relações e as práticas (DELANEY; BATES, 2015). O programa Foster relacionou os elementos essenciais de aprendizagem de todo o ciclo de pesquisa, incluindo desde os conceitos de Ciência aberta até as práticas de gestão de dados com todos os públicos-alvo e as partes interessadas como os estudantes de doutorado, os pesquisadores, os gestores de projetos de pesquisa, os gerentes de conhecimento e bibliotecários e as agências de fomento (FACILITATE OPEN SCIENCE TRAINING FOR EUROPEAN RESEARCH, 2015).

Nesse sentido, várias bibliotecas acadêmicas e de pesquisa estão começando a assumir esse papel mediador e mais ativo na gestão de dados em seus *campi*, aplicando os princípios da biblioteconomia para ajudar a resolver os problemas decorrentes do dilúvio de dados (WITT, 2012). Algumas estão adaptando os serviços de instrução e referência para atender diretamente às necessidades de dados, por exemplo, oferecendo letramento em dados e serviços de referência de dados, ajudando os usuários a encontrar dados e integrá-los em seu aprendizado, ensino e pesquisa (WITT, 2012). Dessa forma um novo leque de serviços tem sido ofertado nas bibliotecas acadêmicas para dar suporte ao pesquisador em um ambiente de pesquisa intensivo em dados, dos quais o letramento em dados é um dos mais recorrentes.

3.3 LETRAMENTO EM DADOS – CONCEITOS E PRÁTICAS

Pelo fato do letramento informacional em dados estar em um período embrionário (CARLSON; JOHNSTON, 2015), a terminologia neste campo ainda não está padronizada (KOLTAY; HORSTMANN; WITT, 2016) e a conceituação passa inicialmente por uma discussão sobre nomenclaturas. Além do uso do termo letramento em dados, há o termo “letramento informacional em dados” adotado por Carlson *et al.*, (2011, p. 2, tradução nossa), “letramento em ciência de dados” usado por Qin e D’Ignazio (2010, p. 2, tradução nossa), e “letramento em dados de pesquisa”, escolhido por Schneider (2013, p. 135, tradução nossa). Koltay (2015a) defende que a terminologia deveria ser unificada em torno do termo “letramento em dados” (KOLTAY, 2015a, p. 403, tradução nossa).

A presente pesquisa, no entanto, adota os conceitos de letramento informacional em dados de Carlson *et al.* (2011), pois estes autores apresentam diversos letramentos ou

alfabetizações informacionais necessárias no contexto de pesquisa intensiva em dados, distinguido, mais detalhadamente dentre elas, o “letramento informacional em dados” de “letramento em dados”, “letramento estatístico” e “letramento informacional” com relação à produção e ao consumo de informações. Agrupa, dessa forma, as várias formas de letramento ao “unir os conceitos de pesquisador como produtor e pesquisador como consumidor de dados” (CARLSON *et al.*, 2011, p. 6, tradução nossa).

O conceito de letramento em dados não está separado por uma fronteira rígida do letramento informacional, visto que este último sempre esteve interessado na compreensão e no uso adequado dos dados que são convertidos em informações (SCHNEIDER, 2013). Encontra-se o letramento em dados também na interseção entre a comunicação científica e o letramento informacional (ASSOCIATION OF COLLEGE AND RESEARCH LIBRARIES, 2013).

Carlson *et al.* (2011) situam as competências em letramento em dados dentro dos padrões de competência informacional da ACRL, a fim de propor os componentes essenciais de um programa de letramento informacional em dados administrado por bibliotecários. A concepção mais ampliada de letramento informacional já é bem conhecida na área da biblioteconomia e muitas universidades oferecem instrução em letramento informacional, seja incorporada em todo o currículo ou como uma série de *workshops* por meio dos sistemas de bibliotecas (WANNER, 2015).

A instrução em letramento informacional nas universidades geralmente é orientada pelo Comitê de melhores práticas em letramento informacional da Association of College And Research Libraries (ACRL) (ASSOCIATION OF COLLEGE AND RESEARCH LIBRARIES, 2012a) que produziu dois padrões principais que norteiam essas práticas nas Universidades Associadas:

- 1) os Padrões de Competência de Letramento Informacional para o Ensino Superior, adotados pelo comitê nos anos 2000 (ASSOCIATION OF COLLEGE AND RESEARCH LIBRARIES, 2000), padrão adotado por 15 anos e que foram recentemente complementados pelo *Framework* de Letramento Informacional para o Ensino Superior (ASSOCIATION OF COLLEGE AND RESEARCH LIBRARIES, 2015);
- 2) as características dos Programas de Letramento Informacional que ilustram as Melhores Práticas, revisada pela última vez em 2012 (ACRL INFORMATION LITERACY BEST PRACTICES COMMITTEE, 2012).

conceitua o letramento informacional como "um conjunto de habilidades que exigem que os indivíduos reconheçam quando a informação é necessária e tenham a capacidade de localizar, avaliar e usar efetivamente as informações necessárias".

Para Calzada Prado e Marzal (2013), letramento em dados e letramento informacional pertencem a uma espécie de contínuo, enquanto outros sugerem que o letramento em dados é um componente menor de um conceito maior de letramento da informação. Por exemplo, considerando o letramento em informação de dados à luz do *framework* da ACRL para o letramento informacional, Shorish (2015) argumenta que o letramento em dados deveria ser tratado como qualquer uma das outras competências de alfabetização e incorporado no fluxo de trabalho dos bibliotecários. Porém, ambas as conceituações de letramento em dados e letramento informacional têm aspectos semelhantes pois têm suas raízes no uso e na prática da informação (WANNER, 2015).

Letramento em dados atualmente é uma competência tão essencial quanto o letramento informacional (CALZADA PRADO; MARZAL, 2013). As habilidades literárias acadêmicas e as habilidades de letramento informacional serão cada vez mais vistas como uma das principais competências de pós-graduação (DELANEY; BATES, 2015) e serão ainda mais importantes devido à crescente complexidade do panorama informacional (NEW MEDIA CONSORTIUM, 2013).

Carlson *et al.* (2011) descrevem como os professores expressaram entusiasticamente a importância das habilidades de letramento em dados, mas muitas vezes não conseguiram definir com precisão o que é o letramento e reconheceram que nas suas próprias práticas de gestão de dados elas estavam ausentes (CARLSON *et al.*, 2011).

No estudo de Carlson *et al.* (2011), em muitos casos, o corpo docente expressou que o treinamento de letramento em dados era esporádico, ou que se esperava que os alunos fossem ao laboratório com habilidades em dados já desenvolvidas. Neste estudo, os professores muitas vezes não estavam preparados para ensinar aos seus alunos e assistentes de pesquisa a estrutura conceitual e as ferramentas necessárias para o entendimento do letramento em dados, acreditando no papel dos bibliotecários para apoiar o desenvolvimento dos programas educacionais de letramento em dados, participando ativamente das discussões (CARLSON *et al.*, 2011).

Novas percepções de letramento informacional encontraram a ideia de que o conhecimento é “construído social e individualmente”, e a noção de letramento como uma prática social (DELANEY; BATES, 2015, p. 39, tradução nossa). Lloyd (2010, p. 22 *apud*

DELANEY; BATES, 2015, p. 39)²⁴ sugeriu que “o letramento informacional é uma prática social que facilita o conhecimento sobre os ambientes de informação nos quais uma pessoa está situada”). Portanto, o letramento informacional será alterado de acordo com o contexto de pesquisa.

Wanner (2015) identifica três grandes tendências das quais emerge a importância do letramento em dados:

- i) desenvolver habilidades na próxima geração, para a qual a gestão de dados está se tornando cada vez mais importante;
- ii) usar as melhores práticas na gestão de dados de pesquisa e;
- iii) desenvolver habilidades de pensamento crítico em estudantes universitários, aplicáveis em uma ampla gama de objetivos, incluindo a força de trabalho, a academia e a vida cotidiana.

A abordagem da Association of College and Research Libraries – ACRL (2013), que optou pelo uso do termo letramento em dados (*data literacy*), está na compreensão ou no direcionamento de como encontrar e avaliar dados, enfatizando a versão do conjunto de dados e a pessoa responsável por ele, não negligenciando as questões de citação e do uso ético dos dados. Para a ACRL (2013), o letramento em dados é uma educação concebida para aqueles que usarão os dados e necessitarão de instrução sobre como compreendê-los e interpretá-los.

Mandinach e Gummer (2013, p. 30) definem o letramento em dados como "a capacidade de entender e usar dados efetivamente para informar as decisões". Eles acrescentam que é um conjunto de habilidades específicas e uma base de conhecimento que permite transformar dados em informações e, finalmente, em um conhecimento utilizável. As habilidades do letramento em dados incluem saber como identificar, coletar, organizar, analisar, resumir e priorizar os dados. As duas últimas habilidades são especialmente dignas de atenção, pois são as habilidades que não aparecem em outras conceituações (MANDINACH; GUMMER, 2013).

O conceito de letramento informacional em dados, de Carlson *et al.* (2011), foi adotado nessa pesquisa pois funde os conceitos de letramento necessários aos pesquisadores enquanto produtores e consumidores de produtos de dados, baseando-se na integração de outros letramentos (dados científicos, estatística e informação) como um conjunto de habilidades emergentes cujos atributos estão baseados em **habilidades funcionais** de coleta, processamento, gestão, avaliação e uso de dados (CARLSON *et al.*, 2011).

²⁴ LLOYD, Annemaree. **Information Literacy Landscapes: Information Literacy in Education, Workplace and Everyday Contexts**. Oxford: Chandos Publishing, 2010.

Igualmente, o letramento em dados permite aos indivíduos acessar, interpretar, avaliar criticamente, gerenciar, manipular e usar dados de forma ética. A gestão, tal como aparece nesta conceituação, compreende a preservação e a curadoria dos dados (CALZADA PRADO; MARZAL, 2013). Calzada Prado e Marzal (2013) também enfatizam a importância de saber como selecionar e sintetizar dados e combiná-los com outras fontes de informação e conhecimento prévio. Eles também enumeram as seguintes habilidades como principais:

- compreender os dados;
- encontrar e obter os dados;
- ler, interpretar e avaliar os dados;
- gerenciar os dados;
- usar os dados.

A compreensão dos dados está ligada ao contexto em que esses dados foram gerados. A dissociação de dados de seu contexto e a perda de contexto torna a reutilização difícil ou impossível (SCHNEIDER, 2013). A condição para a compreensão do contexto dos dados está na forma de documentação e dos metadados, exigência para que os dados sejam reutilizáveis (GIARLO, 2013). Entretanto, um estudo mostrou que a falta de habilidades básicas de gestão de dados, desde o nível de graduação e que se estende aos programas de pós-graduação, é que os alunos tinham uma falta geral de compreensão justamente da documentação e da organização dos dados (RILEY; TREVOR, 2016). Porém, não se limitando a habilidade de documentação, o letramento em dados está na capacidade de desenvolver hipóteses, de identificar problemas, de interpretar os dados, de determinar, de planejar, de implementar, bem como de monitorar cursos de ação, também habilidades necessárias que estão entre os requisitos para adaptar a letramento em dados para seus usos específicos (KOLTAY; SPIRANEC; KARVALICS, 2016).

Um dos estudos mais amplos sobre a gestão de dados de pesquisa cita a falta de treinamento profissional como principal deficiência na gestão efetiva dos dados de pesquisa no longo prazo (COUNCIL ON LIBRARY AND INFORMATION RESOURCES, 2013). Diversas iniciativas de letramento em dados têm se consolidado especialmente nos países que estão sujeitos às políticas mandatórias de depósito de dados de pesquisa. O treinamento em fontes de dados tem sido oferecido em aulas específicas ou como parte de programas de letramento informacional em dados (WHITMIRE, 2015). Whitmire (2015) descreve um programa de letramento em dados para a pós-graduação, já que muitos estudantes de pós-

graduação provavelmente não estão preparados para depositar um conjunto de dados no Repositório Institucional, porque eles não receberam treinamento formal em gestão de dados durante seus programas e, como tal, seus dados não estarão em condições suficientes para serem compartilhados (ou seja, bem organizados e possivelmente documentados). Se preservar e compartilhar os conjuntos de dados produzidos por estudantes de pós-graduação fosse um objetivo compartilhado entre as escolas de Pós-Graduação e da biblioteca universitária, então deveriam ser desenvolvidos mecanismos que capacitassem os estudantes/pesquisadores para alcançar esse objetivo (WHITMIRE, 2015).

3.4 LETRAMENTO EM DADOS COMO UM SERVIÇO DAS BIBLIOTECAS ACADÊMICAS

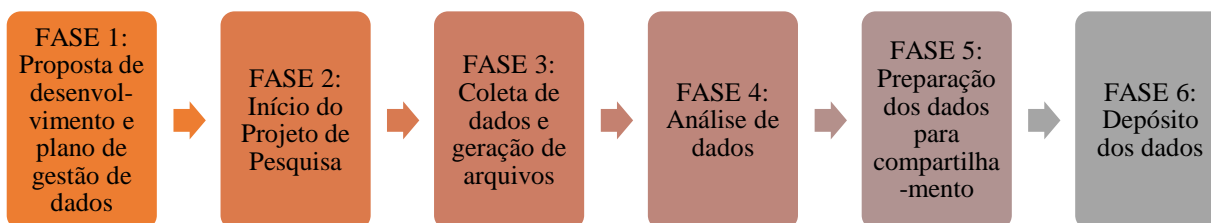
Os serviços de dados de pesquisa (Research Data Services – RDSs) consistem em um amplo espectro de serviços informativos e técnicos que uma biblioteca oferece aos pesquisadores na gestão do ciclo de vida completo dos dados (TENOPIR; BIRCH; ALLARD, 2012). Para que a gestão desses dados seja executada, diversos modelos de ciclo de vida dos dados têm sido propostos.

Os modelos analisados nesta pesquisa foram propostos por instituições de destaque que advogam pelo acesso aos dados dentro da própria comunidade científica, tornando-se referência para pesquisadores e gestores de dados (BALL, 2012), dentre eles, destacam-se: Digital Curation Centre (DCC) Curation Lifecycle Model, DataONE Data Lifecycle, DDI Combined Lifecycle Model, UK Data Archive Data Lifecycle. Para este estudo selecionou-se o modelo de ciclo de vida Data Life Cycle do Inter-university Consortium for Political and Social Research (ICPSR), que enfatiza as melhores práticas da gestão de dados por meio do ciclo de vida dos dados, trabalhando muito próximo aos pesquisadores que submetem seus conjuntos de dados para o uso da comunidade de pesquisa em Ciências Sociais (INTER-UNIVERSITY CONSORTIUM FOR POLITICAL AND SOCIAL RESEARCH, 2012).

O arquivamento de dados é um processo que acompanha todo o ciclo de vida do estudo que estará sendo realizado, devendo ser um elemento incorporado à própria metodologia da pesquisa (INTER-UNIVERSITY CONSORTIUM FOR POLITICAL AND SOCIAL RESEARCH, 2012). As seis etapas do ciclo de vida proposto pela ICPSR são apresentadas na Figura 9, iniciando pela FASE 1: Proposta de desenvolvimento e plano de gestão de dados; FASE 2: Início do Projeto de Pesquisa, FASE 3: Coleta de dados e geração de arquivos; FASE 4: Análise de dados; FASE 5: Preparação dos dados para compartilhamento; FASE 6: Depósito

dos dados.

Figura 9 - Etapas da gestão dos dados propostas pela inter-University Consortium for Political and Social Research (ICPSR)



Fonte: Adaptado de ICPSR (2012, p. 9).

A sequência da gestão dos dados que constam na Figura 9 representa as etapas que o pesquisador precisará executar para que os seus dados de pesquisa sejam disponibilizados. Cada uma dessas etapas carrega inúmeros desdobramentos como regras de conformidade, práticas de gestão, protocolos, formatos e requisitos legais para que os dados estejam disponíveis não apenas de forma aberta, mas também com segurança, com qualidade atestada e preservada frente à obsolescência tecnológica. Desta forma, os dados de pesquisa deixam de figurar como subprodutos das atividades de pesquisa e se tornam um foco de grande interesse para todo o mundo científico (SAYÃO; SALES, 2016).

3.4.1 Características dos Research Data Services (RDSs)

Os bibliotecários têm discutido, historicamente, o presente e o futuro dos Research Data Services (RDSs) ou serviços de dados de pesquisa e suas possíveis funções integradas ao ciclo de gestão de dados (ASSOCIATION OF RESEARCH LIBRARIES, 2006; ASSOCIATION OF RESEARCH LIBRARIES, 2010; COUNCIL ON LIBRARY AND INFORMATION RESOURCES, 2008; TENOPIR *et al.*, 2013).

Tenopir *et al.* (2016; 2017) atestam que bibliotecas universitárias podem desempenhar um importante papel de liderança na prestação de RDSs à comunidade acadêmica e no estabelecimento de uma cultura e de infraestrutura para os RDSs (TENOPIR *et al.*, 2015b).

Para Delaney e Bates (2015), as bibliotecas não apenas oferecem o suporte, mas elas também são colaboradoras da pesquisa científica. A biblioteca deve ser desenvolvida como um polo de serviços informacionais em dados e o bibliotecário deve ter o conhecimento sobre o assunto e a pesquisa. Delaney e Bates (2015) descrevem as seguintes responsabilidades da

biblioteca acadêmica para que continue sendo relevante ao pesquisador:

- a) desenvolver o conhecimento e a confiança do time de trabalho da biblioteca: a equipe da biblioteca tem a responsabilidade profissional de atualizar seus próprios conhecimentos sobre gestão de dados e outras questões relativas à pesquisa;
- b) fornecer orientações ao pesquisador de dados; instruir sobre acesso aberto e outras comunicações acadêmicas; incluir o pessoal da biblioteca acadêmica antes do início dos projetos de pesquisa, na fase de proposta;
- c) desenvolver a conscientização do pesquisador quanto aos seus dados;
- d) ofertar o letramento em dados aos estudantes de pesquisa de pós-graduação, desenvolvendo uma compreensão de como futuros pesquisadores irão gerar e usar dados;
- e) trazer o letramento em dados para o aprendizado da pesquisa de graduação, gerando oportunidade para desenvolver boas práticas.

Os RDSs têm sido ofertados nas bibliotecas acadêmicas em vários países e podem ser caracterizados sob diversos aspectos, como:

- a) oferta de serviços por níveis de formação, como por exemplo para a graduação (MOONEY *et al.*, 2014; WITT *et al.*, 2017), e para a pós-graduação (ADAMICK; REZNIK-ZELLEN; SHERIDAN, 2012; FONG; WANG, 2015);
- b) oferta de serviços por formatos de treinamento: seminários ou *workshops* gerais ou específicos por disciplinas (JOHNSON; BRESNAHAN, 2015; MASSACHUSETTS INSTITUTE OF TECHNOLOGY, 2014; UNIVERSITY OF MINNESOTA, 2014), tutoriais *online* e *e-learning* (JEFFRYES; JOHNSTON, 2013), bem como cursos isolados (CARLSON *et al.*, 2011; QIN; D'IGNAZIO, 2010);
- c) oferta de serviços em universidades de diferentes portes: em universidades de grande expressão (ADAMICK; REZNIK-ZELLEN; SHERIDAN, 2012; TENOPIR; BIRCH; ALLARD, 2012, WITT, 2012) como em unidades menores (FONG; WANG, 2015);
- d) oferta de serviços direcionados às várias áreas do conhecimento, por exemplo: nas Ciências da Terra e do Meio Ambiente (FONG; WANG, 2015); na área de Negócios e na Saúde Pública (MACY; COATES, 2016); na pesquisa Biomédica (FEDERER *et al.*, 2016); na Física (FONG; WANG, 2015); na Meteorologia (FRANK; PHARO, 2016), na Educação (MANDINACH; GUMMER, 2013); nas Ciências Sociais (DECHMAN; SYMS, 2014), nos estudos do Clima (ELLWEIN *et al.*, 2014); na Medicina (SURKIS *et al.*, 2017); na

Geografia (HELBIG, 2016) e nas Ciências Humanas (BINGERT; BUDDENBOHM, 2016). Como um único modelo não serve para todos, o letramento adaptado aos alunos de um departamento específico ou mesmo de um grupo de pesquisa é mais eficaz (FONG; WANG, 2015). As habilidades necessárias de letramento em dados também são consideradas específicas por disciplinas (CARLSON; JOHNSTON, 2015).

Os RDSs têm sido exigidos a partir do atual paradigma de pesquisa, intensivo em dados, e pode ser integrado aos serviços da biblioteca acadêmica (TENOPPIR *et al.*, 2014). Esses serviços auxiliam pesquisadores com manutenção de dados para uso, compartilhamento e reutilização de longo prazo (ELSEVIER, 2013). Os benefícios do uso desses serviços são a maior possibilidade de novas inferências, visibilidade e impacto na pesquisa por meio dos dados, informações e resultados da pesquisa dos cientistas (ELSEVIER, 2013). Verbaan e Cox (2014) mostraram que a biblioteca é a única organização que tem reivindicado esse campo como sua jurisdição. Isso pode ser visto como uma extensão de sua atuação existente em acesso aberto e no tradicional letramento informacional (VERBAAN; COX, 2014).

Os RDSs são altamente complexos, pois incluem gestão de dados de pesquisa (Research Data Management - RDM) e curadoria de dados, governança de dados, letramento em dados, qualidade de dados e padronização das citações de dados (KOLTAY, 2017b). Para Koltay (2017b), os RDSs têm como principais atividades:

- gestão de dados;
- curadoria e administração de dados;
- letramento em dados.

A gestão de dados abrange atividades que não estão especificamente ligadas à biblioteca, mas que podem ser realizadas por ela. Inclui cuidar dos dados de pesquisa, facilitando seu acesso, preservando e agregando valor a esses dados ao longo de seu ciclo de vida (UNIVERSITY OF EDINBURGH, 2015). As bibliotecas podem prestar serviços em dados, documentação e citação, planejamento de gestão de dados, propriedade intelectual e direitos autorais, privacidade e confidencialidade. Eles podem oferecer ajuda sobre hospedagem de dados, repositório seleção e preservação (KOLTAY, 2017b).

A curadoria de dados vai além da gestão de dados, pois compreende serviços adicionais para preservar e agregar valor dentro do ciclo de vida do projeto de pesquisa e além, ou seja, permitindo a reutilização (KOLTAY, 2017b). A administração de dados inclui atividades que

visam manter a integridade e a preservação ao acesso aos dados, onde são armazenados e preservados, como são descritos, como podem ser descobertos, acessados e reutilizados (KOLTAY, 2017b).

3.4.2 Estudo dos Tipos de RDSs

O letramento em dados constitui o quadro principal dentre os tipos de RDSs informativos ou de referência, dentre uma vasta gama de serviços das bibliotecas no apoio à pesquisa intensiva em dados (TENOPIR *et al.*, 2014), dentre eles:

- a) consulta aos professores, aos funcionários ou aos alunos em planos de gestão de dados (DMP);
- b) consulta aos professores, aos funcionários ou aos alunos em padrões de dados e de metadados;
- c) divulgação e colaboração com outros provedores RDSs, dentro ou fora do *campus*;
- d) fornecimento de suporte de referência para encontrar e citar dados ou conjuntos de dados;
- e) criação de instruções *online* e ajuda de busca para dados, conjuntos de dados ou repositórios de dados;
- f) discussão de RDSs com outros bibliotecários, ou outras pessoas no *campus*, ou Profissionais de RDSs.

O letramento em dados pode ser útil em todos os processos em que os dados são gerenciados, sendo vital para realizar os RDSs (KOLTAY, 2015b). O letramento em dados compartilha uma série de recursos com vários outros letramentos, especialmente o letramento informacional, principalmente porque enfatiza as perspectivas críticas do seu objeto (KOLTAY, 2015b; KOLTAY, 2017b).

Para Koltay (2017b), a avaliação da qualidade é uma das atividades mais importantes relacionadas com o letramento em dados. Em Calzada Prado e Marzal, (2013), essa relação entre letramento e qualidade dos dados é vital. Ela inclui, por exemplo, a avaliação de fontes de dados para confiabilidade e para evitar erros ou problemas. A avaliação já é importante para quem coleta os dados, mas a importância da análise crítica é crucial na interpretação de dados de outros para que se apresente conclusões corretas. A qualidade dos dados também está intimamente ligada à governança de dados, que não recebeu atenção considerável na área da Ciência da Informação (KOLTAY; HORSTMANN; WITT, 2016), mas é uma das condições

necessárias para a viabilidade de dados abertos e traz conhecimento substancial ao trabalho do bibliotecário acadêmico (KOLTAY; HORSTMANN; WITT, 2016). De acordo com Giarlo (2013), a qualidade dos dados é determinada pelos seguintes fatores principais: (i) confiança, (ii) autenticidade, (iii) segurança, (iv) compreensibilidade, (v) descoberta, (vi) acessibilidade e (vii) integridade.

A governança de dados refere-se a órgãos institucionais, responsáveis pelas regras, políticas, padrões; direitos de decisão; responsabilidades e métodos de execução, fornecendo diretrizes sobre disponibilidade e possibilidades de acesso, proveniência, confiabilidade e significado dos dados (KOLTAY, 2017b).

Uma pesquisa foi realizada em 2011 sobre os RDSs como parte integrante das responsabilidades de trabalho dos bibliotecários de instituições vinculados à Association of Research Libraries – ARL. A ARL é uma organização sem fins lucrativos de 125 bibliotecas de pesquisa em instituições de pesquisa nos EUA e no Canadá (ASSOCIATION OF RESEARCH LIBRARIES – ARL, [s.d.]). Quase três quartos dos bibliotecários responderam à pesquisa e, apesar de que na época a frequência desses serviços era baixa, já havia evidências de que muitos bibliotecários da ARL acreditavam ter o conhecimento, as habilidades e as oportunidades para fornecer RDSs no futuro, motivados por interesses profissionais, por demanda dos usuários e por responsabilidades inerentes ao seu trabalho (TENOPIR *et al.*, 2013). O estudo de Tenopir *et al.* (2013) mostrou que os bibliotecários acreditavam que os RDSs eram importantes para serem oferecidos pelas bibliotecas acadêmicas, pois são serviços relacionados com a missão e o papel das bibliotecas e que contribuíssem para o aumento da visibilidade e o impacto da pesquisa institucional (TENOPIR *et al.*, 2013).

Em 2011 e 2014, foi pesquisada uma amostra de bibliotecas acadêmicas norte-americanas que são membros da Association of College and Research Libraries (ACRL) para determinar quais RDSs eram oferecidos e quais se pretendiam oferecer no futuro (TENOPIR; BIRCH; ALLARD, 2012; TENOPIR *et al.*, 2015b, TENOPIR *et al.*, 2016).

Na pesquisa de 2011 identificou-se que a maioria das bibliotecas da amostra ainda não ofereciam RDSs, porém estavam planejando fazê-lo no futuro (TENOPIR; BIRCH; ALLARD, 2012). A pesquisa de acompanhamento encontrou pouca mudança no percentual de bibliotecas oferecendo RDSs (TENOPIR *et al.*, 2015b). Entrevistas subsequentes, com diretores de bibliotecas, identificaram que as instituições de maior porte ofereciam os RDSs com maior recorrência do que as instituições menores. Entre as que ofereciam, eram mais frequentes os serviços de consulta ou de referência em dados do que os serviços técnicos ou aplicados (TENOPIR *et al.*, 2017). Ambas as pesquisas identificaram que as instituições de maior porte

ofereciam os RDSs com maior recorrência do que as instituições menores. Entre as que ofereciam, eram mais frequentes os serviços de consulta ou de referência em dados do que os serviços técnicos ou aplicados (TENOPIR *et al.*, 2017).

Em 2016, as pesquisas anteriores da ACRL foram replicadas em bibliotecas de pesquisas acadêmicas europeias que são membros da Ligue des Bibliothèques Européennes de Recherche (Liber) – (Associação das Bibliotecas de pesquisa europeias) (TENOPIR *et al.*, 2017). A pesquisa identificou as práticas dos tipos de serviços de apoio à gestão de dados oferecidos por essas instituições, já que diversos países da Europa foram os primeiros a requererem dados de pesquisa (TENOPIR *et al.*, 2017). Dos 333 membros acadêmicos da Liber, a maioria dos 119 respondentes indicou que já estavam oferecendo serviços RDSs, além do que diversas outras também estavam planejando fazê-lo futuramente. Os diretores dessas bibliotecas reconheceram a importância de fornecer os RDSs para suas comunidades (TENOPIR *et al.*, 2016; TENOPIR *et al.*, 2017). Dentre os serviços ofertados, estava previsto a capacitação dos pesquisadores em gestão de dados, entretanto, esta oferta dependia de quanto os profissionais dessas bibliotecas estariam aptos para esse tipo de suporte (TENOPIR *et al.*, 2017).

A maioria das bibliotecas europeias (TENOPIR *et al.*, 2017), assim como as bibliotecas norte-americanas investigadas em pesquisas anteriores, também ofereciam ou planejavam oferecer serviços mais consultivos do que técnicos (TENOPIR; BIRCH; ALLARD, 2012; TENOPIR *et al.*, 2015b). Para Tenopir *et al.* (2016), isso pode ser explicado porque os serviços consultivos são mais alinhados com os serviços de referência existentes nas bibliotecas e os conjuntos de habilidades para os RDSs técnicos exige níveis de habilidades adicionais, maior dedicação de tempo e de recursos por parte da biblioteca e seus funcionários. No entanto, foi apenas recentemente que as bibliotecas acadêmicas reconheceram a necessidade de ter funcionários com habilidades em gestão de dados de pesquisa (MAYERNIK *et al.*, 2014; TENOPIR, 2016). Para Carlson *et al.* (2011), o letramento em dados pode ser direcionado para a educação de profissionais de gestão de dados bem como para pesquisadores, podendo ter uma finalidade dupla (CARLSON *et al.*, 2011).

As bibliotecas acadêmicas europeias oferecem oportunidades para que os membros da equipe desenvolvam habilidades em RDSs, com a maioria dessas oportunidades sob a forma de apoio à equipe para participar de atividades de treinamento e grupos fora da biblioteca (TENOPIR *et al.*, 2016). Os diretores das bibliotecas concordam firmemente sobre a importância dos RDSs, bem como sua relevância para o papel tradicional da biblioteca como administrador de todos os tipos de produção acadêmica. Esse papel também foi atribuído à

biblioteca na pesquisa dos diretores de bibliotecas da América do Norte, já que fornecer tecnologia e serviços para apoiar a preservação e gestão de dados de pesquisa é uma extensão natural desse papel em uma época em que tanto a quantidade de dados de pesquisa quanto às políticas e leis em torno da gestão de dados estão aumentando (TENOPIR *et al.*, 2016).

A fim de oferecer plenamente RDSs técnicos, as bibliotecas precisam ter pessoal tecnologicamente qualificado ou aumentar consideravelmente as oportunidades de treinamento em tecnologia para seus funcionários, o que pode não ser viável devido a restrições de recursos (TENOPIR *et al.*, 2015b). Existem muitas habilidades relacionadas aos dados que os bibliotecários, e as equipes das bibliotecas, precisam adquirir para acompanhar a evolução do cenário de pesquisa e, frise-se, os custos relacionados ao desenvolvimento da força de trabalho são significativos (TENOPIR *et al.*, 2015b). No âmbito europeu várias bibliotecas que apoiam o RDSs oferecem oportunidades para o desenvolvimento dessas habilidades (TENOPIR *et al.*, 2017). Quase 84% das bibliotecas afirmaram ter oportunidades para que a equipe desenvolvesse habilidades em RDSs (TENOPIR *et al.*, 2017).

Tenopir, Birch e Allard (2012) evidenciam que a maioria das bibliotecas pesquisadas no âmbito da ARL está atribuindo às equipes internas o trabalho com a gestão de dados ao invés de contratar novos profissionais e que a preparação dessas equipes pode ser na prática do trabalho ou mais formalizada (TENOPIR *et al.*, 2013). A mesma tendência tem ocorrido na Europa. As recomendações finais do relatório do grupo de trabalho em EScience / Research Data Management da Liber recomenda fortemente a questão da requalificação, evidenciando a sua importância já que poucas bibliotecas atualmente têm condições de contratar pessoal novo e especializado. A biblioteconomia de dados é uma profissão em si, mas muito trabalho em serviços de dados pode ser feito por profissionais da informação requalificados (CHRISTENSEN-DALSGAARD *et al.*, 2012).

Programas de formação em Ciência da Informação (Library and Information Studies – LIS) também estão desenvolvendo cursos e programas para formar os profissionais da informação com uma base sólida na gestão e na conservação de dados (ASSOCIATION OF COLLEGE AND RESEARCH LIBRARIES, 2012b; KARON *et al.*, 2013; TENOPIR; BIRCH; ALLARD, 2012; VARVEL; BAMMERLIN; PALMER, 2012).

Para Kafel, Creamer e Martin (2014), os bibliotecários sem formação científica ou que não estejam familiarizados com os ambientes de pesquisa científica (por exemplo, pessoal de pesquisa, projeto, fluxos de trabalho, instrumentação, protocolos) estão preocupados com sua competência para fornecer instrução de gestão de dados de pesquisa aos seus pesquisadores (KAFEL; CREAMER; MARTIN, 2014).

De acordo com Heidorn (2011), os dados em instituições de pesquisa e ensino são muito heterogêneos, sendo difícil para qualquer pessoa ter todas as habilidades necessárias para fornecer aos pesquisadores em sua instituição com RDSs (HEIDORN, 2011). Por isso, uma abordagem colaborativa pode ser a melhor solução para implementar RDSs (TENOPIR *et al.*, 2015b). Implementar RDSs em uma biblioteca acadêmica requer recursos humanos, financeiros, de tempo, qualificação e apoio institucional. Uma solução é o trabalho associativo entre as bibliotecas com outras universidades para que compartilhem recursos (TENOPIR *et al.*, 2015b) A colaboração com grupos fora da biblioteca ou com outras instituições seria essencial para o desenvolvimento completo de RDSs para uma comunidade acadêmica.

Hey Tansley e Tolle (2009) também recomendam a atuação de forma colaborativa entre as partes interessadas para viabilizar o desenvolvimento da infraestrutura tecnológica necessária para suportar os RDSs e a gestão dos dados em longo prazo (HEY; TANSLEY; TOLLE, 2009). Nesse contexto, Tenopir *et al.* (2014) e Koltay (2017a) atribuem ao letramento em dados a qualidade de um dos pilares dentre os tipos de RDSs ofertados, cujos resultados têm sido um instrumento importante na colaboração com a pesquisa acadêmica orientada a dados (KOLTAY, 2017a; TENOPIR *et al.*, 2014).

A realização de pesquisas intensivas em dados em universidades e centros de pesquisa dependerá cada vez mais das habilidades e das competências de um grupo profissional relativamente novo e sofisticado (JAHNKE; ASHER; KERALIS, 2012). Responder a esses desafios requer uma força de trabalho científica capacitada e diversificada para gerar novos conhecimentos.

Igualmente importantes são profissionais aptos para ofertar o letramento em dados de pesquisa, capazes de resolver problemas em um mundo cada vez mais complexo e baseado em dados (KARON *et al.*, 2013). Tanto nos Estados Unidos quanto no exterior, as escolas da Biblioteca e da Ciência da Informação estão cada vez mais preocupadas com o desenvolvimento de currículos para atender a crescente necessidade de profissionais bem treinados para prestar serviços e ferramentas para realizar curadoria, compartilhar, preservar e reusar os dados de pesquisa científica (KARON *et al.*, 2013). Enquanto alguns centros de pesquisa e universidades desenvolveram sistemas e serviços para apoiar esses objetivos, a maioria dos próprios pesquisadores tem pouco ou nenhum letramento em práticas de gestão de dados (JAHNKE; ASHER; KERALIS, 2012).

Vários autores desenvolveram estudos que mapearam as competências²⁵ necessárias e as competências essenciais para o letramento em dados de pesquisa como detalhado na próxima seção. Esse mapeamento é apresentado em formato de *frameworks* ou estruturas de competências identificadas tanto na literatura como em programas de letramento em dados existentes.

3.5 FRAMEWORKS DE LETRAMENTO EM DADOS

Tanto autores como Carlson *et al.* (2011), Jeffryes e Johnston (2013), Calzada Prado e Marzal (2013), Carlson e Johnston (2015), quanto instituições como Vitae (2011) e Foster (2015) interessadas nos dados de pesquisa propuseram modelos de competências (*frameworks*) necessários para o letramento em dados para o ensino superior e para o *eScience*. Esses *frameworks* foram desenvolvidos a partir de distintas metodologias. Por exemplo, Carlson *et al.* (2011), assim como Jeffryes e Johnston (2013), realizaram levantamentos de necessidade junto aos estudantes e pesquisadores, por meio de questionários, para basear os seus modelos propostos de *frameworks*. Calzada Prado e Marzal (2013) descreveram seu *framework* a partir da análise da literatura. Já Carlson e Johnston (2015) criaram o projeto Data Information Literacy – DIL a partir de um projeto colaborativo disciplinar. Outro *framework* foi elaborado a partir de critérios de qualidade estabelecidos para pesquisadores de alto nível (VITAE, 2011). O *framework* disponível no ambiente virtual Facilitate Open Science Training For European Research (FOSTER, 2015) visa o cumprimento de políticas de financiamento para atender os requisitos do Horizon 2020.

Os seis *frameworks* propostos nessa seção (CALZADA PRADO; MARZAL, 2013; CARLSON; JOHNSTON, 2015; CARLSON *et al.*, 2011; FOSTER, 2015; JEFFRYES; JOHNSTON, 2013; VITAE, 2011) estão detalhados no Quadro 3.

²⁵ Durand (1998 *apud* DURAND, 2000, conceitua a competência sob três aspectos interdependentes: o conhecimento (informações registradas pelo indivíduo), as habilidades (colocar o conhecimento em prática) e as atitudes (o querer fazer) necessários para atingir um objetivo.

Quadro 3 - Apresentação comparativa dos *frameworks* das competências do letramento em dados

Frameworks das competências do letramento em dados	Base/Nome do Projeto	Metodologia	Referência
<i>Framework 01</i>	Baseado no <i>framework</i> de letramento informacional da ACRL (2000)	Levantamentos de dados empíricos entre estudantes e pesquisadores por meio de questionários	Carlson <i>et al.</i> (2011). Quadro 34
<i>Framework 02</i>	Baseado no <i>framework</i> de letramento informacional da ACRL (2000)	Análise bibliográfica	Calzada Prado; Marzal (2013) Quadro 35
<i>Framework 03</i>	Facilitate Open Science Training For European Research	Atendimento das políticas de financiamento Horizon 2020	Foster (2015) Quadro 36
<i>Framework 04</i>	Vitae Researcher Development Framework -	Análise qualitativa de critérios para pesquisadores de alto nível	Vitae (2011) Quadro 37
<i>Framework 05</i>	Data Information Literacy – DIL	Projeto colaborativo institucional	Carlson; Johnston (2015); Quadro 38
<i>Framework 06</i>	University of Minnesota Data Management Course	Projeto colaborativo institucional	Jeffryes; Johnston (2013) Quadro 39

Fonte: Autoria própria (2018).

Todos os seis *frameworks* apresentam pontos de interesse para este estudo. Carlson *et al.*, (2011), por exemplo, estabeleceram 11 competências essenciais²⁶ de dados com base em suas pesquisas ao interagir com professores sobre práticas de dados. O trabalho de Carlson *et al.*, (2011) foi usado como fundamento para o desenvolvimento do projeto Data Information Literacy – DIL (CARLSON; JOHNSTON, 2015). Essas onze competências essenciais incluem habilidades de letramento em dados que podem ser ensinadas por bibliotecários; elas são detalhadas no Apêndice G – Quadro 38:

- 1) introdução aos bancos de dados e formatos de dados;
- 2) descoberta e aquisição de dados;
- 3) gestão e organização de dados;
- 4) conversão de dados e interoperabilidade;
- 5) metadados;
- 6) curadoria de dados e reutilização;
- 7) culturas da prática;

²⁶ Prahalad e Hamel (1990) diferenciam algumas competências, destacando-as como "competências essenciais" (*core competences*). Essas são as competências que geram vantagens, são distintas e únicas, difíceis de serem imitadas e que geram valor percebido, sendo imprescindíveis para o desenvolvimento de determinada ação.

- 8) preservação de dados;
- 9) análise de dados;
- 10) visualização de dados;
- 11) ética, incluindo citação de dados.

Calzada Prado e Marzal (2013) instituíram um *framework* das competências e das habilidades do letramento em dados de acordo com as semelhanças encontradas entre as competências listadas na literatura, incluindo também quaisquer outras que possam ser considerados apropriadas. O *framework* de Calzada Prado e Marzal (2013) é detalhado no Apêndice G – Quadro 35 – e em linhas gerais estabelece as seguintes ações:

- entender dados;
- encontrar e / ou obter dados;
- ler, interpretar e avaliar os dados;
- gerenciar dados;
- usar dados.

Com objetivo de apoiar a integração das melhores práticas da Ciência Aberta, oferecendo suporte às diferentes partes interessadas no processo de comunicação científica, especialmente jovens pesquisadores, em adotar o acesso aberto no contexto do Espaço Europeu da Investigação e no cumprimento das políticas de acesso aberto e regras de participação prevista para o Horizonte 2020 (H2020), desenvolveu-se o projeto Facilitate Open Science Training For European Research²⁷ – Foster (CARVALHO; RODRIGUES; PRÍNCIPE, 2014).

Uma das iniciativas do projeto é oferecer o suporte com conteúdo de treinamento relevante, disponibilizado no Portal Foster em módulos de autoaprendizagem e *e-learning* que acompanham o plano de aprendizagem. O plano está estruturado em um nível crescente de competências, buscando a integração do aprendizado na rotina diária de pesquisa (FACILITATE OPEN SCIENCE TRAINING FOR EUROPEAN RESEARCH, 2015). O

²⁷ São parceiros do projeto: Universidade do Minho - Portugal; Georg-August-Universitaet Goettingen – Germany; Danmarks Tekniske Universitet – Denmark; Stichting eIFL.net – Netherlands; SPARC-Europe – UK; Stichting LIBER – Netherlands; University of Glasgow – DCC – UK; Technische Universiteit Delft – Netherlands; The Open University – UK; ICM - Uniwersytet Warszawski – Poland; Consortium Universitaire de Publications Numériques Couperin – France; Consejo Superior de Investigaciones Científicas – Spain; University of Edinburgh - DCC – UK, além de outras 27 organizações de 13 países (instituições, escolas, associações de estudantes e jovens investigadores, etc.), comprometeram-se a colaborar e apoiar o projeto Foster. (CARVALHO; RODRIGUES; PRÍNCIPE, 2014).

framework do plano de ensino descreve os objetivos de aprendizagem efetivos e é estruturado por tópicos de Ciência Aberta, que acompanham as principais responsabilidades das partes interessadas da pesquisa (Estudantes de Pós-Graduação, Pesquisadores, Gerentes de Projetos de Pesquisa, Gerentes de Conhecimento e Bibliotecários, Administradores Institucionais e agências de financiamento) ao longo do Ciclo de Vida da Pesquisa (FACILITATE OPEN SCIENCE TRAINING FOR EUROPEAN RESEARCH, 2015). Este é apresentado no Apêndice G – Quadro 36, e, em especial, adota os requisitos para:

- Ciência Aberta;
- gestão de dados de pesquisa (Research Data Management – RDM);
- questões legais;
- ética.

O Vitae Researcher Development Framework (RDF) (VITAE, 2011), *framework* de desenvolvimento de pesquisadores da Vitae, uma entidade sem fins lucrativos do Reino Unido, incorporou habilidades de gestão de dados como um dos elementos da pesquisa bem-sucedida. Este *framework* é estruturado em quatro domínios (VITAE, 2011) que cobrem:

- 1) o domínio dos conhecimentos e as habilidades intelectuais: este domínio contém o conhecimento e as habilidades intelectuais necessárias para realizar excelentes pesquisas;
- 2) o domínio da eficácia pessoal: este domínio contém as qualidades pessoais, habilidades de carreira e autogerenciamento necessárias para apropriar-se e controlar o desenvolvimento profissional;
- 3) o domínio da governança e da organização da pesquisa: este domínio contém o conhecimento dos padrões, requisitos e conduta profissional necessária para o gerenciamento efetivo da pesquisa;
- 4) o domínio do engajamento, influência e impacto: este domínio contém o conhecimento, a compreensão e as habilidades necessárias para se envolver, influenciar e impactar no contexto acadêmico, social, cultural e econômico.

O Apêndice G – Quadro 37 – apresenta o *framework* com os descritores que indicam os conhecimentos, os comportamentos e os atributos de um pesquisador bem sucedido, mapeados pelo Vitae (2011) e que tiveram os dados de pesquisa incorporados.

Um projeto Interorganizacional entre as bibliotecas da Universidade de Purdue, em parceria com as bibliotecas da Universidade de Minnesota, da Universidade de Oregon e da Cornell University, tinha como um dos objetivos elaborar um *framework* em letramento em dados por meio do desenvolvimento e implementação de programas de instrução de letramento informacional para estudantes de pós-graduação, de forma a permitir que os alunos aprendessem habilidades apropriadas ao seu contexto disciplinar (CARLSON; JOHNSTON, 2015). O projeto Data information Literacy (DIL) foi resultado de cinco estudos de casos nas áreas de Engenharia Elétrica e Computação, Engenharia Biológica e Agrícola, Recursos Naturais, Engenharia Civil, Arquitetura e Ecologia / Paisagem das instituições envolvidas e teve a duração de dois anos (CARLSON; JOHNSTON, 2015). As cinco equipes realizaram uma análise coletiva de entrevistas e intervenções educativas para identificar padrões e semelhanças entre as experiências em desenvolvimento de programas DIL, bem como para quaisquer diferenças significativas. Na maioria das competências identificadas, a falta de treinamento formal para adquirir habilidades importantes surgiu como um tema comum entre as disciplinas (CARLSON; JOHNSTON, 2015). O *framework* com as competências e habilidades identificadas no projeto Data information Literacy (DIL) estão listadas no Quadro 39 do Apêndice G, sendo as principais:

- práticas culturais;
- conversão de dados e interoperabilidade;
- curadoria e reutilização de dados;
- gestão e organização de dados;
- preservação de dados;
- processamento e análise de dados;
- qualidade de dados e documentação;
- visualização de dados e representação;
- bancos de dados e formatos de dados;
- descoberta e aquisição de dados;
- ética e atribuição de autoria;
- metadados e descrição de dados.

A proposta de letramento de Jeffryes e Johnston (2013) baseada em seus modelos de *frameworks* consta no Apêndice G – Quadro 39 – e apresenta as seguintes habilidades ou

necessidades de aprendizagem dos estudantes em relação ao gerenciamento de dados:

- introdução a dados de gestão;
- dados a serem gerenciados;
- organização e documentação de métodos;
- acesso a dados e a propriedades;
- compartilhamento de dados e sua reutilização;
- técnicas de preservação;
- execução do seu Data Management Plan – DMP.

Alguns destes quadros conceituais são baseados previamente em modelos genéricos de letramento informacional (CARLSON *et al.*, 2011; CALZADA PRADO; MARZAL, 2013) ou propõem competências genéricas que podem ser aplicáveis a quaisquer áreas do conhecimento (FACILITATE OPEN SCIENCE TRAINING FOR EUROPEAN RESEARCH, 2015; VITAE, 2011). Porém, diante das especificidades e características inerentes a cada campo da Ciência (SAYÃO; SALES, 2014) o letramento em dados precisa estar contextualizado ao ambiente disciplinar (MAYBEE; ZILINSKI, 2015).

3.5.1 *Frameworks* Baseados em Modelos Genéricos *versus* Especializados

Apesar de alguns autores (CALZADA PRADO; MARZAL, 2013; CARLSON *et al.*, 2011) considerarem o *framework* de letramento informacional da ACRL (2000), como um modelo que tem uma abrangência generalista, para o “desenvolvimento de um conjunto de competências em dados”, Maybee e Zilinski (2015, p. 1) alertam para o desafio de se estabelecer modelos genéricos para a habilidade de pessoas em usarem informações em diferentes contextos (MAYBEE; ZILINSKI, 2015). Salienta-se que contextos diferentes exigem competências específicas, sendo um desafio para atendimento. Os *frameworks* de letramento em dados estariam sujeitos aos mesmos desafios. Como proposta de avanço de um *framework* em letramento em dados genérico, Maybee e Zilinski (2015) apresentam a abordagem de *informed learning framework*, que se traduziu por letramento contextual, já que considera aprendizado em como usar os dados no contexto disciplinar para elaboração dos *frameworks* de letramento em dados (MAYBEE; ZILINSKI, 2015).

Este processo compreende três etapas: (i) analisar os *frameworks* existentes de letramento em dados para identificar aspectos-chave; (ii) identificar os principais aspectos dos *frameworks* em que o letramento em dados é considerado como um elemento dentro de contextos disciplinares; (iii) adaptar um modelo de letramento em dados selecionado para desenvolver um novo *framework* capaz de ensinar os alunos do ensino superior a se envolverem com dados dentro de contextos de aprendizagem disciplinar (MAYBEE; ZILINSKI, 2015).

Analisando alguns *frameworks* de letramento em dados, Maybee e Zilinski (2015) apontaram algumas delimitações de competências em dados que podem ser vistas como alinhadas com uma abordagem genérica do letramento em dados, entre elas: consciência sobre os dados e seu papel, acesso aos dados, avaliação, análise, organização e interpretação de dados existentes; tomada de decisões com base em dados, gerenciamento dos dados, incluindo organização e análise, desenvolvimento de protocolos para segurança e armazenamento, compartilhamento e documentação, comunicação, síntese, visualização e representação de dados; uso ético e reconhecimento das fontes de dados, avaliação e gerenciamento de riscos, entendimento dos problemas relacionados à reutilização de dados e à sua preservação, cuidado com as práticas de curadoria para armazenamento e uso em longo prazo.

Alguns autores como Bruce *et al.* (2013) sugerem, porém, que o aprendizado de habilidades genéricas não prepara as pessoas, por extensão, para usar informações e dados, em contextos da vida real. Portanto, novos *frameworks* devem ser desenvolvidos para reformular o letramento e os quadros conceituais adequados às demandas de contextos disciplinares, o que resultaria em *frameworks* em letramento contextual de dados, ou seja, quadros conceituais de aprendizados necessários para o uso de dados por parte do pesquisador, que sejam aderentes às características da sua área de conhecimento.

As práticas da abordagem letramento contextual são extraídas de áreas acadêmicas, disciplinares ou profissionais relevantes para o que está sendo estudado (MAYBEE; ZILINSKI, 2015). O letramento contextual envolve a intersecção de várias áreas de especialização, incluindo especialistas nas disciplinas, tecnólogos e bibliotecários, e, portanto, sugere-se que todas essas partes devem se unir para desenvolver ações sob essa abordagem. As estratégias de letramento contextual são baseadas em solução de problemas ou aprendizagem guiada por questionamentos que tendem a envolver informações para aprender sobre um assunto disciplinar (MAYBEE; ZILINSKI, 2015).

Maybee e Zilinski (2015) propõem, portanto, com base na ideia de letramento contextual, uma abordagem para letramento contextual de dados, como uma alternativa ao letramento em dados genérico, que desloca o foco para a aquisição de habilidades relacionadas

aos ambientes contextuais, para que os estudantes aprendam como usar dados em suas áreas de estudos disciplinares (BRUCE; HUGHES, 2010; BRUCE *et al.*, 2013). Ellwein *et al.*, (2014) também apresentam o letramento contextual como uma abordagem que intensifica o engajamento no processo de letramento em dados.

Três aspectos da abordagem letramento contextual de dados são benéficos para o desenvolvimento de um *framework* de letramento em dados em contextos de aprendizagem disciplinar: (i) orientação do conteúdo do letramento relacionado aos dados (o tipo e o uso dos dados são determinados com base em como os estudantes obtém os resultados de aprendizagem focados nas suas disciplinas); (ii) incentivo a relevância do letramento (os resultados relacionados ao uso dos dados são determinados pelo contexto disciplinar); e (iii) apoio a aprendizagem de longo prazo (utilizar dados no contexto disciplinar e da preparação dos alunos para usar os dados no contexto profissional) (MAYBEE; ZILINSKI, 2015). As características do letramento contextual (BRUCE; HUGHES; SOMERVILLE, 2012), adaptadas aos dados, sugerem maneiras de educar os alunos do ensino superior a usar dados que se concentram na sua disciplina, o que abrangeria as várias maneiras pelas quais os dados são usados nos contextos acadêmicos, disciplinares ou profissionais, o que incluiria não apenas a pesquisa, mas também a tomada de decisões baseadas em dados (MAYBEE; ZILINSKI, 2015).

O projeto desenvolvido a partir do *framework* de Carlson e Johnston (2015), Projeto Data Information Literacy (DIL Project), apesar de ser inspirado na estrutura mais genérica proposta pela Association of College and Research Libraries (ACRL) para o letramento informacional, tem seu *framework* estritamente alinhado com os conceitos de letramento contextual (CARLSON; JOHNSTON, 2015). O *framework* foi construído a partir de entrevistas e de programas educacionais em diferentes contextos disciplinares. Porém, os autores Carlson e Johnston (2015) reconhecem que haverá necessidade de bibliotecários criarem variáveis específicas das competências e de habilidades em letramento em dados que podem estar relacionadas tanto com as práticas e as necessidades disciplinares, como também poderiam ser baseadas em um método particular de pesquisa, de tipo de dados ou de contexto específico.

Esses entendimentos ratificam a presente pesquisa na proposta conceitual de um ambiente virtual para o letramento em reuso de dados contextualizadas nas disciplinas das Ciências Sociais.

Além dos quadros conceituais, outra variável desta pesquisa são os recursos de aprendizagem *online*, disponíveis nos ambientes virtuais de aprendizagem (AVAs) e que buscam promover o letramento para a gestão dos dados de pesquisa. Em seguida são

apresentados alguns projetos de letramento informacional em dados que tem sido desenvolvido de forma contextualizada às áreas do conhecimento que pretendem cobrir.

3.6 RECURSOS EM AMBIENTES VIRTUAIS DE APRENDIZAGEM

Para Pereira, Schmitt e Dias (2007), os ambientes virtuais de aprendizagem (AVAs) podem ser conceituados como mídias²⁸ utilizadas no espaço virtual para veicular conteúdo e permitir a interação entre os atores do processo educativo. Essas mídias intermediam o processo de ensino-aprendizagem à distância, também conhecida como Educação à distância (EaD). Outras terminologias também são empregadas, como: aprendizagem baseada na internet, educação ou aprendizagem *online*, ensino ou educação à distância via internet. No idioma inglês as referências registram os termos *web-based learning*, *online learning*, *learning management systems*, *virtual learning environment* e *e-learning* (PEREIRA; SCHMITT; DIAS, 2007, p. 5).

A popularização da internet, e os avanços científicos e tecnológicos somados à demanda social, permitiu a construção de AVAs, viabilizando a comunicação simultânea entre os participantes separados geograficamente a qualquer hora. A escolha do AVA como modalidade de ensino permite que se rompam barreiras temporais e geográficas de acesso à informação (PEREIRA; SCHMITT; DIAS, 2007).

Quanto ao processo educativo, Mattar (2014) salienta que o ambiente virtual transcende a incorporação da tecnologia em si e que a aprendizagem se dá nos movimentos de integração e de colaboração inerentes a esses ambientes. Os movimentos são caracterizados pelas interações dos alunos com o conteúdo, da observação de outros interagindo com o professor, dos alunos entre si, da própria ferramenta, do ambiente e da interface²⁹ (MATTAR, 2014). Os modelos de *designs* dos objetivos de aprendizagem, das atividades e das abordagens dos ambientes também podem ser direcionados por estas variadas formas de interação (MATTAR, 2014). Mattar (2014) propõe que o *design* de interações e de atividades, enquanto práticas críticas do *design* educacional, incluam pelo menos os seguintes elementos-chave: Objetivos; Tipos, Nível e Estrutura das Atividades (Síncrona ou Assíncrona); Recursos; Mídias; Ferramentas; Avaliação e Duração (MATTAR, 2014).

O processo de elaboração de conteúdo de um AVA envolve um trabalho

²⁸ As mídias sob a perspectiva do AVA referem-se ao “conjunto de ferramentas e recursos tecnológicos resultantes da evolução das tecnologias de comunicação e informação que permitem a emissão e a recepção de mensagens” (PEREIRA, SCHMITT; DIAS, 2007, p. 4).

²⁹ Tecnologia usada na apresentação do conteúdo (MATTAR, 2014).

multidisciplinar, cujos profissionais de diversas áreas do conhecimento trabalham de forma integrada buscando garantir a coerência entre as estratégias de ensino e o seu produto final (PEREIRA; SCHMITT; DIAS, 2007). É preciso uma gama de profissionais envolvidos, exercendo funções como a de conteudista, de *designer* instrucional, de assessores linguísticos, de *designer* gráfico, de programador, de *web*-roteirista, de *web-designer* e de ilustrador/animador (PEREIRA; SCHMITT; DIAS, 2007).

A ênfase da atual pesquisa, contudo, está pautada nos esforços para a construção de uma proposta de conteúdo em detrimento das funcionalidades e técnicas da EaD. Para Pereira, Schmitt e Dias (2007), essas funcionalidades são componentes que organizam conteúdos, controlam acessos e tempo, acompanham atividades, oferecem suporte *online*, avaliam a apreensão da aprendizagem e permitem a comunicação eletrônica (PEREIRA; SCHMITT; DIAS, 2007).

As bibliotecas de pesquisa precisam de informações confiáveis sobre as necessidades de dados dos pesquisadores para desenvolver serviços que apoiem os pesquisadores (KAFEL; CREAMER; MARTIN, 2014). Diante dessa necessidade, diversas iniciativas de bibliotecas acadêmicas desenvolveram recursos *online* para o letramento em dados (KAFEL; CREAMER; MARTIN, 2014).

Diversos recursos em ambientes virtuais de aprendizagem estão disponíveis. Alguns exemplos de maior expressão são o DataONE, o Projeto Data Information Literacy (DIL Project), o New England Collaborative Data Management Curriculum (NECDMC), o MANTRA *Research Data and Management Training*. Alguns detalhamentos:

- o projeto DataONE (Data Observation Network for Earth) Education Modules, conta com módulos de letramento voltados para as Ciências Ambientais (DATAONE, 2012);
- o Projeto Data Information Literacy (DIL Project), promove o Curso de Gestão de Dados e é destinado a estudantes de várias Engenharias. O projeto Data Information Literacy Project é uma colaboração das bibliotecas da Universidade Purdue, da Universidade Cornell, da Universidade de Minnesota e da Universidade de Oregon, como já citado. O objetivo do projeto é desenvolver uma infraestrutura em que os alunos aprendam habilidades de letramento em dados relevantes para sua disciplina enquanto auxilia a desenvolver um processo no qual os currículos da DIL possam ser articulados nas comunidades de pesquisa (RILEY; TREVOR, 2016). Os resultados mais relevantes até agora incluem estudos de caso em cada universidade e um curso de gestão de dados. Esses recursos estão disponíveis *online* e ajudam os pesquisadores a criar os planos de gestão de dados. O Institute of Museum and Library Science financiou este projeto e os materiais

do curso também estão disponíveis gratuitamente sob uma licença Creative Commons CC-BY (RILEY; TREVOR, 2016);

- o New England Collaborative Data Management Curriculum (NECDMC) de autoria de uma coalizão de bibliotecários aborda a necessidade de materiais educacionais para que os bibliotecários possam usar para ensinar a gestão de dados de pesquisa a diversas disciplinas STEM (sigla em inglês para as disciplinas de Ciência, Tecnologia, Engenharia e Matemática) (KAFEL; CREAMER; MARTIN, 2014);
- o NECDMC compreende sete módulos que se alinham com as recomendações do plano de gestão de dados da NSF. Aborda desafios que enfrentam pesquisadores nas ciências e seu currículo usa estudos de caso para contexto e até mesmo oferece documentação sobre como ensinar cursos usando o material fornecido. A National Library of Medicine (Biblioteca Nacional de Medicina dos EUA) financia o projeto, que está disponível gratuitamente para uso sob uma licença Creative Commons (CC-BY) (RILEY; TREVOR, 2016);
- o MANTRA Research Data and Management Training (UNIVERSITY OF EDINBURGH; EDINA Data Library, 2017), que consiste em oito tópicos principais, cada um projetado para levar cerca de uma hora, incluindo questionários interativos, vídeos de pesquisadores que descrevem como lidaram com problemas de gestão de dados e leitura adicional. Uma seção adicional sobre o tratamento de dados amplia as habilidades para trabalhar dentro de quatro ambientes específicos de *software* (UNIVERSITY OF EDINBURGH, EDINA Data Library, 2017).

Outros exemplos de recursos e seus objetivos são apresentados no Quadro 4. Destaca-se que o público previsto do ambiente de pesquisa, para Carlson *et al.*, (2011), são os estudantes e os pesquisadores acadêmicos. Carlson e Johnston (2015) consideram os docentes e os pesquisadores, assim como os estudantes da pós-graduação e do pós-doutorado como os usuários do ambiente de letramento.

Já o *Framework* do Facilitate Open Science Training For European Research (2015) relaciona o público alvo ao longo de todo o ciclo da pesquisa, incluindo estudantes de Pós-Graduação, pesquisadores, gerentes de projetos de pesquisa, gerentes de conhecimento e bibliotecários, administradores institucionais e agências de financiamento. Neste estudo, o público majoritário será o de estudantes, de pesquisadores e de docentes com interesse na área das Ciências Sociais.

Esgotadas, na perspectiva deste estudo, as observações da Revisão de Literatura, é

apresentado, a seguir, o Alinhamento Conceitual entre os dados de pesquisa e o letramento, sob as lentes da teoria ator-rede.

Quadro 4 - Recursos para o letramento em dados disponíveis em ambientes virtuais de aprendizagem

(continua)

Recurso/Projeto	Descrição	Público Alvo	Link
01- Bodleian Data Library – University of Oxford (UNIVERSITY OF OXFORD, 2018-2019).	Foco em habilidades gerais e recursos, que podem ser usados em uma ampla gama de disciplinas. Os <i>workshops</i> abrangem tópicos desde a gestão de seus dados de pesquisa até a descoberta de informações.	Pós-graduados e pesquisadores	https://libguides.bodleian.ox.ac.uk/workshops/rdm
02- Consortium of European Social Science Data Archives (CESSDA) e European Research Infrastructure Consortium (ERIC). (CONSORTIUM OF EUROPEAN SOCIAL SCIENCE DATA ARCHIVES, 2018).	Oferece treinamento em gestão de dados de pesquisa, curadoria de dados, etc. nas Ciências Sociais. Reúne os arquivos de dados das Ciências Sociais em toda a Europa, com o objetivo de promover os resultados da pesquisa em ciências sociais e apoiar a pesquisa e a cooperação nacional e internacional.	Pesquisadores de Ciências Sociais que estão em um estágio inicial de gestão de dados de pesquisa, os provedores também oferecem cursos de treinamento para funcionários, bibliotecários e pesquisadores.	https://www.cessda.eu/Research-Infrastructure/Training/Expert-Tour-Guide-on-Data-Management
03- Data Management Course (JOHNSTON; JEFFRYES, 2012).	Espaço de trabalho para o projeto de pesquisa financiado pelo Institute of Museum e Library Services (IMLS) que investiga as necessidades de alfabetização informacional de dados (DIL). Este projeto, composto por equipes de pesquisa da Universidade de Purdue, da Universidade de Minnesota, da Universidade de Oregon e da Universidade de Cornell.	Estudantes e corpo docente	http://www.datainfolit.org/Data-Management-Course-The-University-of-Minnesota https://sites.google.com/a/umn.edu/data-management-course_structures/home-1
04- Data Observation Network for Earth – (DataONE) (DATAONE, 2012).	Curso de curta duração de 2 dias realizado em Santa Barbara, desenvolvido pelo Grupo de Trabalho DataONE Community Education and Engagement (CEE)	Três estudantes de mestrado, 10 estudantes de doutorado e 2 pesquisadores de pós-doutorado participaram do <i>workshop</i> .	www.dataone.org/education-modules .

(continua)

Recurso/Projeto	Descrição	Público Alvo	Link
05- Digital Curation Center - DC 101 materials: Access and Reuse (DIGITAL CURATION CENTER, 2014-2019).	<p>O DC 101 foi desenvolvido porque o DCC, em seu papel de fonte de orientação especializada e orientação para a comunidade, identificou a necessidade de uma introdução contextual e teórica dos fundamentos da curadoria digital com exemplos práticos e exercícios. O curso indica o que deve ser considerado no planejamento e na implementação de projetos. Resultado de uma parceria entre a University of Edinburgh, a University of Glasgow, UKOLN (based at the University of Bath) e o Science and Technology Facilities Council (STFC).</p>	<p>O público-alvo são novos afiliados no DCC. Módulos: Access and reuse. Modulo acesso e reuso e um guia para citação de conjuntos de dados, brevemente como compartilhar dados médicos e como usar Creative Commons.</p>	<p>http://www.dcc.ac.uk/training/train-the-trainer/dc-101-training-materials</p>
06- Essentials 4 Data Support (RESEARCH DATA NETHERLANDS, <i>online</i>).	<p>Research Data Netherlands oferece um curso introdutório aos conhecimentos básicos e habilidades (essenciais) para permitir que um facilitador em letramento de dados dê os primeiros passos para apoiar pesquisadores no armazenamento, gestão, arquivamento e compartilhamento de seus dados de pesquisa.</p>	<p>Colaboradores de dados (dados) bibliotecários, equipe de TI e pesquisadores</p>	<p>http://datasupport.researchdata.nl/en/</p>
07- Facilitate Open Science Training for European Research (FOSTER, 2015).	<p>O Foster portal é uma plataforma de <i>e-learning</i> que reúne os melhores recursos de treinamento dirigidos àqueles que precisam saber mais sobre o Open Science</p>	<p>Desde pesquisadores em início de carreira, a gerentes de dados, bibliotecários, administradores de pesquisa e escolas de pós-graduação</p>	<p>www.fosteropenscience.eu/; Open Data Use and Reuse +60 recursos</p>
08- Managing and Sharing Research Data handbook (CORTI <i>et al.</i> , 2014).	<p>Página de acompanhamento do livro do UK Data Service Gerenciando e Compartilhando Dados de pesquisa, com apresentações relacionadas e exercícios em grupo.</p>	<p>Estudantes da pós-graduação</p>	<p>www.ukdataservice.ac.uk/manage-data/handbook</p>
09- Research Data and Management Training - MANTRA (UNIVERSITY OF EDINBURGH, 2017).	<p>O MANTRA é um curso de oito unidades <i>online</i> com diretrizes para ajudá-lo a entender e refletir sobre como gerenciar os dados digitais coletados em toda a sua pesquisa. Em colaboração com o Institute for Academic Development como parte do projeto patrocinado pela Jisc - Managing Research Data project (2010-11).</p>	<p>Estudantes de pós-graduação, nos programas da University of Edinburgh nos campos de Geociências, Ciências Sociais e Políticas e Psicologia Clínica.</p>	<p>https://mantra.edina.ac.uk/</p>

(conclusão)

Recurso/Projeto	Descrição	Público Alvo	Link
10- New England Collaborative Data Management Curriculum (NECDMC) (NECDMC, 2012).	<p>É uma ferramenta instrucional para ensinar práticas recomendadas de gestão de dados nas disciplinas de ciências da saúde, ciências e engenharia. Alinha-se com as recomendações do plano de gestão de dados da National Science Foundation e aborda desafios universais de gestão de dados. É liderado pela Biblioteca Lamar Soutter da Escola de Medicina da Universidade de Massachusetts em parceria com várias bibliotecas da região da Nova Inglaterra.</p>	Universitários, estudantes de pós-graduação e pesquisadores	https://library.umassmed.edu/resources/necdmc/index
11- Research Data Management and Sharing - Coursera (CURATING RESEARCH ASSETS AND DATA USING LIFECYCLE EDUCATION (CRADLE); UNIVERSITY OF NORTH CAROLINA; UNIVERSITY OF EDINBURGH, 2017).	<p>Fornece uma introdução à gestão e compartilhamento de dados de pesquisa. Depois de concluir este curso, os alunos entenderão a diversidade de dados e suas necessidades de gestão durante o ciclo de vida dos dados de pesquisa, serão capazes de identificar os componentes de bons planos de gestão de dados e estarão familiarizados com as melhores práticas para trabalhar com dados, incluindo organização, documentação e armazenamento e segurança de dados. Os alunos também compreenderão a importância do arquivamento e compartilhamento de dados, bem como avaliar a confiabilidade dos repositórios. Há um certificado de conclusão está disponível por US\$ 29.00. Desenvolvido Curating Research Assets and Data Using Lifecycle Education (CRADLE), em colaboração com a Universidade da Carolina do Norte em Chapel Hill e com EDINA em Universidade de Edimburgo</p>	Pesquisadores e bibliotecários	www.coursera.org/learn/research-data-management-and-sharing

Fonte: Autoria própria (2019).

4 ALINHAMENTO CONCEITUAL E PROPOSTA INICIAL

O presente capítulo demonstra o viés conceitual desta pesquisa sob o prisma das redes e dos atores no âmbito da Comunicação Científica, buscando a fundamentação teórica e aproximando as concepções de rede para a comunicação científica dos trabalhos de Roosendaal e Geurts (1997) e da abordagem construtivista da Teoria Ator-Rede, em especial de Latour (1998; 2001). Para Latour (2001, p. 218), o conhecimento é construído a partir da “interação de trocas de propriedades de um coletivo de humanos e não humanos”. Portanto, propõe-se nesta pesquisa estabelecer a comunicação científica como uma rede construída a partir das inter-relações de elementos heterogêneos (humanos e não humanos) que ocorre em um ambiente complexo (LATOURE, 1998; 2001). Desta forma, esta proposta corrobora a perspectiva de Roosendaal e Geurts (1997), em que a comunicação científica deixa de figurar como um processo linear.

Neste capítulo discutem-se as formulações dos resultados de pesquisa em torno da caixa preta (LATOURE, 2000; LATOURE; CALLON, 1981), evidenciando na seção 4.1 a disponibilidade dos dados de pesquisa como um actante capaz de permitir o escrutínio da construção dos fatos científicos.

A seção 4.2 discorre sobre o ambiente virtual da rede de comunicação científica e as forças presentes na rede, apoiado em cinco constructos: (i) Recursos Tecnológicos, (ii) e-*Science*, (iii) Comunicação Científica, (iv) Regulação e (v) Ciência Aberta. Também evidencia os relacionamentos entre os distintos elementos inerentes à cada constructo e a como essas relações moldam e projetam os dados, emergindo o letramento para seu efetivo reuso.

Portanto, a construção do objeto de pesquisa se dá sob a ótica da dinâmica desses diversos atores, em que os dados são em si um actante central e como esses estão interagindo com outros atores.

Finaliza-se com a seção 4.3 em que a proposta conceitual da pesquisa é apresentada na forma de requisitos de letramento informacional para o reuso de dados a partir da análise em triangulação entre os *frameworks*, os conjunto de dados reutilizados e os recursos de aprendizado *online* que estejam aderentes às melhores práticas da W3C.

4.1 CONCEITO DE CAIXA PRETA E A COMUNICAÇÃO CIENTÍFICA

Latour (2000), ao abordar a construção científica, denunciou a vulgarização da ciência perante a maioria de leigos na sociedade, mantendo a atividade científica em uma caixa preta em que apenas os estudos sociais de Ciência, Tecnologia e Sociedade (CTS), vinham tentando abrir (LATOURE, 2000, p. 33-34). “A caixa preta contém aquilo que já não precisa ser repensado, coisas cujo conteúdo tornou-se uma questão de indiferença” (LATOURE; CALLON, 1981, p. 284). Ao invés de analisar uma caixa preta fechada, ou o resultado final, como fatos científicos ou artefatos técnicos, Latour (2000) preferia seguir os passos dos cientistas no momento e nos lugares em que os seus elementos constituintes ainda estão instáveis, durante a construção.

Grosso modo, os resultados de pesquisas, publicados no formato de artigos científicos, de teses e dissertações, podem ser comparados a essa caixa preta. São consolidados a partir de teorias, dados e métodos, encapsulados no texto científico. Enquanto a comunicação científica flui da informalidade para a formalidade (BORGMAN, 2007, p. 66), o artigo publicado revisado por pares se constitui um “ponto fixo” ou uma “publicação definitiva” (FRANKEL *et al.*, 2000, p.254). Ainda que haja muitos debates a respeito de onde estaria a fronteira para este ponto fixo em um mundo em rede (BORGMAN, 2007, p. 66), ao considerar a publicação formal uma caixa preta, infere-se que a maneira de **abrir-la seria a partir dos dados de pesquisa disponíveis, gerados pelos trabalhos dos cientistas**, antes que se cristalizem nos resultados publicados de suas análises. Diante dessa disponibilidade, refazer a análise dos dados sob outros enfoques pode maximizar os recursos empregados na coleta desses dados, seja por uma nova análise dos mesmos ou por ratificar estudos já elaborados. Pode ser feita a verificação da caixa-preta a um custo aceitável. A importância da disponibilidade dos dados de pesquisa se dá principalmente pela possibilidade de compreender a dinâmica de geração daquele conhecimento, sem obscuridades.

O processo de investigação que gera dados, no entanto, é muito mais difícil de ser replicado, já que envolve diversos recursos (LATOURE, 2000, p. 115). Latour (2000) descreve as implicações e as circunstâncias da construção de fatos científicos e de artefatos ao ponto de se tornarem caixas pretas, ou seja, o momento em que esses recursos científicos estão acabados, em que toda a sua complexidade está encapsulada em um conceito ou em um artefato, de forma que se torna inquestionavelmente aceito. As caixas-pretas são o último estágio da construção científica. Quaisquer questionamentos ao objeto estabilizado custam muito, discordar e argumentar, abrir a caixa-preta é, para Latour (2000, p. 116), uma tarefa cara, quicá inexecutável,

como já antecipado. Os questionamentos e posicionamentos exigiriam refazer todo o caminho, desde o acesso aos textos, aos pesquisadores, aos institutos de pesquisa que representam os governos e as políticas públicas em que estes institutos são regulados, também aos laboratórios, às matérias primas, às pesquisas de campo, aos dados levantados, aos ambientes naturais ou sociais de coleta desses dados, às habilidades dos cientistas na condução dos seus experimentos e, até mesmo, ao viés filosófico de condução da pesquisa, visto que, segundo Feenberg (2010) a ciência não é neutra.

Para Latour (2000), uma quantidade ínfima de leitores faria a reexecução do que foi feito pelos autores, pois é uma opção cara, que seria inviabilizada, também, pela falta de acesso aos recursos que estão fora dos textos, em especial aos dados da pesquisa, ou aos “recursos que provêm de instrumentos, animais, figuras [...], do lugar de onde saíram: a natureza ou do laboratório” (LATOUR, 2000, p. 101). Refazer esse percurso implicaria na disponibilidade de máquinas e de laboratórios, de teorias e de métodos utilizados. Nessa perspectiva, a disponibilidade de dados de pesquisa ampliaria o reuso dos mesmos não apenas para validações, mas para novas inferências a partir de abordagens distintas, tornando a construção científica transparente, aberta e democrática, corroborando com os estudos de Fecher e Friesike (2014).

Portanto, infere-se que os dados de pesquisa, actantes não humanos, que se configuram como as “inscrições” e que constituem o “alicerce” do texto científico (LATOUR, 2000, p. 101), podem influenciar a comunicação científica. Esta, por sua vez, insere-se em um ambiente interligado, ocorrendo de forma não linear (ROOSENDAAL; GEURTS, 1997), podendo ser caracterizada como uma rede justamente a partir da agência de atores que interagem de forma interdependente, porém interconectada (LAW, 1993). Em seguida busca-se evidenciar a proposta conceitual das forças que moldam o ambiente da rede da comunicação científica, quais os seus constructos e como seus elementos atuam nesta rede.

4.2 O AMBIENTE VIRTUAL DA REDE DE COMUNICAÇÃO CIENTÍFICA

Há mais de duas décadas, Roosendaal e Geurts (1997, p. 23) já identificavam que o conhecimento era um produto virtual e imaterial em si, o que se infere ser, ainda hoje, atual. A comunicação científica desde a década de 1990, principalmente com o advento da internet, já tinha sua forma de disseminação eletrônica (ROOSENDAAL; GEURTS, 1997, p. 23). Isso implicava em canais de distribuição virtuais para viabilizar a divulgação de um produto virtual, utilizando memória virtual. Nessa direção, os departamentos de tecnologia da computação das

instituições criaram alianças com as bibliotecas acadêmicas, que já tinham um papel relevante na disseminação da comunicação científica. Nesse contexto, desde então, a rede de comunicação científica é caracterizada pelo alto nível de interdependência estratégica desses atores – departamentos de tecnologia da computação e bibliotecas acadêmicas de instituições de ensino (ROOSENDAAL; GEURTS, 1997, p. 22).

A rede da comunicação científica transcende então a indústria de informação, não estando limitada ao conteúdo. Nesse sentido, a rede não inclui apenas o compartilhamento da informação depois do processo de pesquisa finalizado, mas promove a troca de todos os tipos de informação durante a própria execução da pesquisa. Fundamentando-se em Roosendaal e Geurts (1997, p.24), a rede da comunicação científica, incluindo todos os atuantes que viabilizam a sua disseminação, se qualifica para uma organização virtual. A consideração da troca de informação durante o processo de pesquisa contribui para reintegração das funções ou tarefas que estão presentes nas partes interessadas “externas”. Dessa forma, a reintegração responde à forte necessidade de novas formas de retorno de investimento do capital de pesquisa (ROOSENDAAL; GEURTS, 1997, p. 24).

Para tal, eles propõem uma estrutura modular da publicação científica. A proposta é garantir o acesso e a utilização dos elementos de informação e viabilizar a integração de módulos com outros artigos ou outros conjuntos de informações e coleções, o que se infere, no contexto atual, que os conjuntos de dados ou *datasets* cumpram esse papel. Esse desenvolvimento requer uma nova organização da rede e uma nova divisão de tarefas e responsabilidades entre os atores. Novas alianças entre os autores serão necessárias, ainda que eles apresentem formas separadas de subgrupos ou subconjuntos dentro da rede. Isso implica na cooperação para a constituição de uma real organização virtual. A citação de dados pode ser um exemplo dessas novas alianças, pois provê a uma comunidade de pesquisa mais ampla a possibilidade de reuso de dados aos pesquisadores de outras áreas de estudos, estimulando a colaboração interdisciplinar (TASK GROUP ON DATA CITATION STANDARDS AND PRACTICES, 2013, p. 42). Além disso, os artigos com dados de pesquisa compartilhada são mais citados, aumentando o seu fator de impacto (PIWOWAR; VISION, 2013, p. 11).

Neste contexto se inserem os ambientes virtuais de aprendizagem (AVA), que têm o potencial de tornar o processo de ensino-aprendizagem mais ativo, dinâmico e personalizado, visto que o espaço virtual é utilizado para promover a interação e a colaboração à distância entre os atores do processo e a interatividade com o conteúdo disponível (PEREIRA; SCHMITT; DIAS, 2007).

Os AVAs, recuperando a acepção de Mattar (2014), não se limitam às tecnologias e sua

incorporação, portanto, mas conjugam as questões de aprendizagem aos movimentos de interação e colaboração. Em especial na elaboração de conteúdos os ambientes virtuais de aprendizagem exigem a atuação de profissionais de diversas áreas do conhecimento (PEREIRA; SCHMITT; DIAS, 2007), para o desenvolvimento de recursos virtuais de aprendizagem, sejam eles os conteúdos propriamente ditos, ou ainda métodos ou tecnologias, sob uma ótica multidisciplinar. Então, deste viés, os ambientes virtuais de aprendizagem incorporam recursos de aprendizagem *online* de uma perspectiva integrada e colaborativa.

A bibliotecas acadêmicas têm sido responsáveis por iniciativas de desenvolvimento de recursos *online* para o letramento em dados, dentro desta abordagem integrada; os currículos dos recursos são pensados em interação com estudantes, pesquisadores e os especialistas técnicos de acordo com Kafel, Creamer e Martins (2014). Exemplos de AVAs para os dados de pesquisa podem ser encontrados no Quadro 4, essas iniciativas listadas reforçam a importância dos AVAs enquanto modelo para o processo de ensino-aprendizagem.

Considerando esse complexo ambiente de comunicação científica constituído de interdependências entre os atores e sua heterogeneidade, a Teoria Ator Rede se constitui um arcabouço teórico e metodológico que viabiliza a identificação desses atores e a sua agência, evidenciando a dinâmica da rede de comunicação científica e as implicações dos dados de pesquisa nesse cenário. Os dados também trazem consigo um requisito novo, a partir do ponto em que são projetados, ou melhor, evidenciados na rede. Há como consequência, a emergência do letramento em dados, que possibilita aos pesquisadores compreender e efetivamente fazer uso desse ativo para novas inferências em suas pesquisas, para citar apenas um benefício.

A estrutura conceitual do ambiente de pesquisa se alicerça sobre cinco constructos para que os dados sejam evidenciados: (i) a comunicação científica, (ii) os aspectos regulatórios, (iii) a Ciência Aberta, (iv) os recursos tecnológicos e (v) o conceito de *eScience*, sendo todos simétricos sob as lentes da Teoria Ator-Rede. Cada um desses constructos influencia o ambiente de pesquisa, exercendo suas forças, como por exemplo:

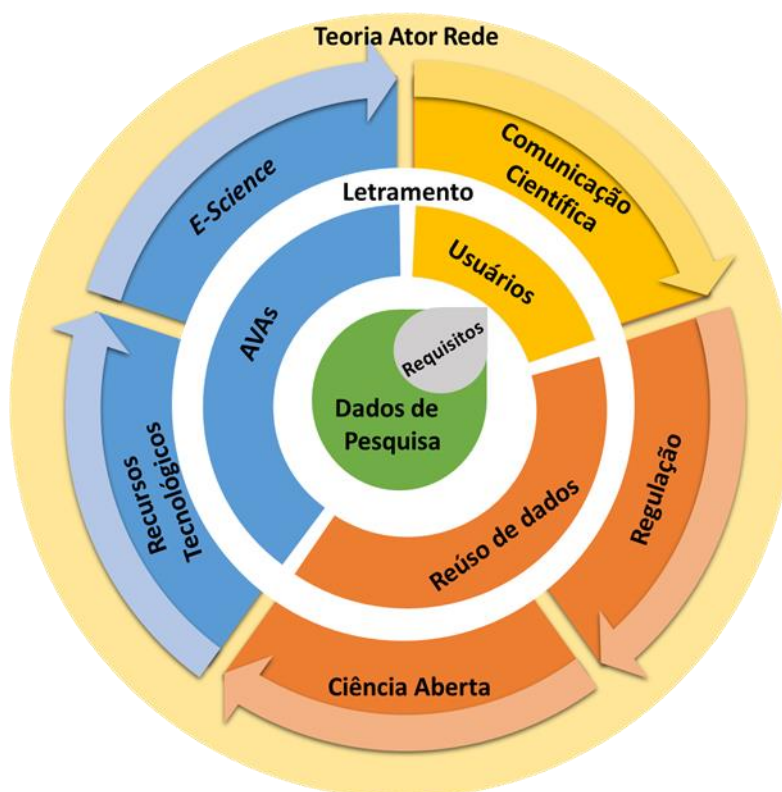
- a rede da comunicação científica, cujos atores estabelecem aspectos regulatórios;
- os aspectos regulatórios que são estimulados por um ambiente cada vez mais movido pelas prerrogativas da Ciência Aberta;
- a Ciência Aberta que estimula o desenvolvimento de uma infraestrutura tecnológica de softwares de códigos fonte abertos e protocolos de acesso aberto que viabilizem o depósito e o acesso aos dados;

- os recursos tecnológicos capazes de suportar o modelo eletrônico de geração de dados do paradigma moderno da *eScience*, (ou a ciência orientada à dados) e da necessidade de disseminação da informação *online*;
- o eScience que por sua vez reconfigura a rede de comunicação científica a partir da projeção dos dados como em ator central na prática da ciência do século XXI.

A ação conjunta dos agentes desses constructos molda a disponibilidade de dados, o que leva à emergência do letramento, considerado em um segundo momento – segunda camada da Figura 10 – a fim de que essa disponibilidade seja útil, viabilizada pelo reuso dos mesmos. Na segunda camada da Figura 10, alguns elementos de maior influência emergem de alguns dos constructos indicados, seguindo o mesmo padrão de cor para identificar o constructo de origem. Destaca-se, portanto, a necessidade de letramento por meio de ambientes virtuais de aprendizagem (AVAs) (elemento que emerge do constructo Recursos Tecnológicos) para que os usuários (atores provenientes do constructo Comunicação Científica) sejam aptos ao reuso de dados (prática inerente ao constructo Ciência Aberta).

No centro, estão os dados de pesquisa, cujos requisitos para o reuso são apenas uma parte das competências necessárias em relação aos mesmos. Não se inclui, portanto, o gerenciamento de dados como um todo, levando-se em conta aspectos como o preparo dos dados para o compartilhamento e a sua gestão ao longo do ciclo da pesquisa. Os AVAs, por sua vez, estão relacionados, sobretudo, com os recursos tecnológicos, pois é a partir desse constructo que são viabilizados. Mas, não apenas isso, os ambientes virtuais de aprendizagem, por sua natureza informacional, rompem as barreiras tecnológicas, temporais e geográficas, conforme bem salientado por Pereira, Schmitt e Dias (2007), aproximando-se do quarto paradigma da ciência (*eScience*) em razão desta ser uma prática intensiva na geração, no uso, no processamento e no compartilhamento dos dados de pesquisa, de acordo com Hey, Tansley e Tolle (2009), de forma colaborativa, em rede e interdisciplinar. Os usuários são parte do constructo da Comunicação Científica, como um dos elementos que formam esta rede (mais detalhada na Figura 11). O reuso de dados é um desdobramento da Ciência Aberta, pois está intrinsecamente ligado aos conceitos de abertura e melhor aproveitamento da atividade científica, ainda que esteja ligado também à Regulação, já que princípios e regras são necessários para que o reuso seja possível. Evidencia-se na Figura 10 como todos esses elementos estão inter-relacionados sob a perspectiva teórica deste trabalho.

Figura 10 - Estrutura conceitual do ambiente de pesquisa



Fonte: Autoria própria (2019).

As relações de interdependência dos elementos de cada um dos constructos são evidenciadas, visto que podem influenciar ou estar diretamente relacionados com elementos ligados ou que sejam inerentes a outro constructo. Essas relações, em que elementos exercem forças ou influências uns aos outros, são o que caracterizam as redes (LAW, 1993). No constructo Recursos Tecnológicos, por exemplo, elementos como Repositórios de dados e Repositórios Institucionais estão diretamente relacionados com os elementos *softwares* de código fonte abertos, visto que essa são estes que viabilizam o desenvolvimento de diversas ferramentas para implantação de repositórios. Os códigos abertos, por sua vez, constituem um elemento relacionado com os valores da Ciência Aberta. Outro exemplo que se relaciona com um elemento da Ciência Aberta é o Ambiente Virtual de Aprendizagem, já que pode ser um dos recursos educacionais utilizadas para o letramento em reúso de dados, como propõe esse trabalho.

Outros exemplos dessas relações estão no constructo *e-Science*, que cria uma maior demanda por Armazenamento em nuvem, viabilizado pela computação em nuvem, sendo esta uma infraestrutura que é elemento do constructo Recursos Tecnológicos. A Abertura e o

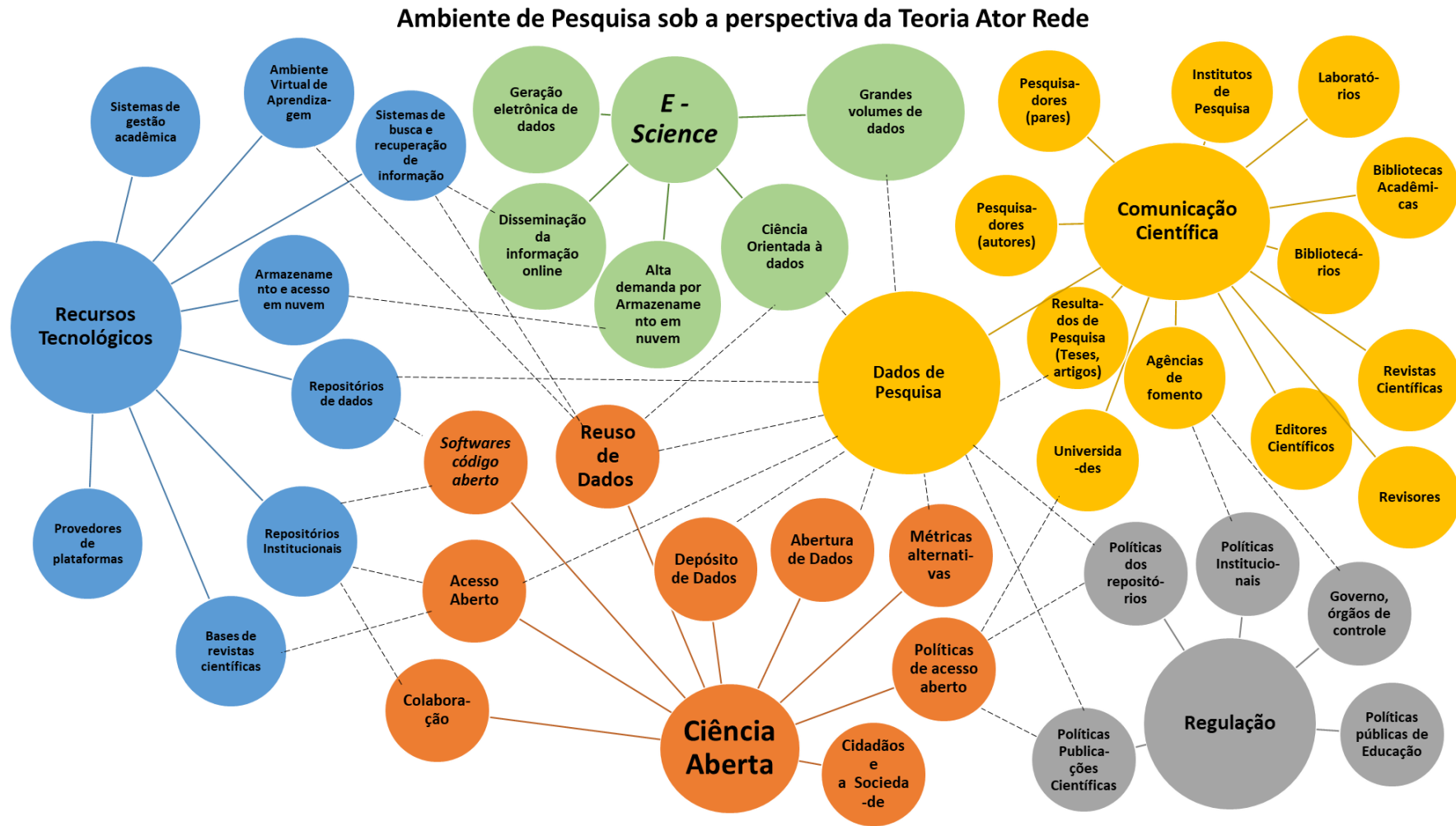
depósito de dados (Ciência Aberta) geraram grandes volumes de dados (*e-Science*) que, para serem disseminados, precisaram estar disponíveis em sistemas de busca e recuperação de informação *online* (também elemento do constructo Recursos Tecnológicos). Políticas (elemento do constructo Regulação) de órgãos de controle, Instituições e Publicações Científicas (elementos do constructo da Comunicação Científica) foram necessárias para regulamentar a abertura dos dados de pesquisa (presente no constructo Ciência Aberta). Portanto, a Figura 11 demonstra os seguintes exemplos destas relações entre elementos de cada um dos constructos.

- a) Constructo Recursos Tecnológicos: Provedores de plataformas, Bases de revistas científicas, Repositórios Institucionais, Repositórios de dados, Sistemas de busca e recuperação de informação, Armazenamento e acesso em nuvem, Sistemas de gestão acadêmica e Ambiente Virtual de Aprendizagem;
- b) Constructo *e-Science*: Ciência Orientada à dados, Disseminação da informação *online*, alta demanda por Armazenamento em nuvem, Geração eletrônica de dados e Grandes volumes de dados;
- c) Constructo Comunicação Científica: Universidades, Agências de fomento, Laboratórios, Bibliotecas Acadêmicas, Bibliotecários, Institutos de Pesquisa, Revistas Científicas, Editores Científicos, Revisores, Resultado de Pesquisa (Teses, Dissertações artigos), Dados de Pesquisa, Pesquisadores e Métricas alternativas;
- d) Constructo Ciência Aberta: *Softwares* de código fonte abertos, Abertura de Dados, Depósito de Dados, Cidadãos e a Sociedade, Políticas de acesso aberto, Acesso Aberto, Abertura de Dados, Colaboração e Métricas alternativas;
- e) Constructo Regulação: Governo, órgãos de controle, Políticas públicas de Educação, Políticas Institucionais, Políticas Publicações Científicas e Políticas dos repositórios;

Sob a perspectiva de rede, os elementos desdobram-se desses constructos e se relacionam entre si e com os dados de pesquisa. Alguns elementos são comuns a mais de um constructo, e essa integração está representada na Figura 11, salientando que a Regulação está destacada por ser tratada como um constructo distinto que, no entanto faz parte do Reúso de Dados.

Sequencialmente apresentam-se as principais teorias norteadoras desta proposta conceitual, organizadas por temas e conceitos que foram apropriados para esta pesquisa.

Figura 11 - Ambiente de pesquisa sob a perspectiva da Teoria Ator Rede



Fonte: Autoria própria (2019).

4.3 PROPOSTA CONCEITUAL DA PESQUISA

Os principais estudos que direcionam a proposta conceitual desta pesquisa estão listados de acordo com o seu tema, a seguir:

- a) a abordagem da Ciência Aberta sob aspectos de Fecher e Friesike (2014), com destaque para as escolas democrática e a pragmática que consideram os dados abertos;
- b) quanto aos benefícios do acesso aberto, apontados por Davis *et al.*, (2008), como o aumento da visibilidade, a melhora no processo de pesquisa e a redução de custos;
- c) quanto a perspectiva da caracterização dos dados de pesquisa, que não está vinculada aos formatos ou às formas de registro, mas está ligada à sua finalidade, segundo o The Engineering and Physical Sciences Research Council (2018, *online*);
- d) quanto à relevância do reúso de dados de pesquisa em Ciências Sociais, esta pesquisa está amparada nos trabalhos de Borgman (2000; 2007), Curty (2016a, CURTY, 2016b, CURTY *et al.*, 2017) e Tenopir (2011, 2015a);
- e) o ciclo de vida da gestão dos dados de pesquisa a ser incorporado à própria metodologia da pesquisa (INTER-UNIVERSITY CONSORTIUM FOR POLITICAL AND SOCIAL RESEARCH, 2012, p.8);
- f) as prerrogativas do letramento em dados estão alinhadas aos preceitos de Koltay (2015a; 2016b; 2017a; 2017b);
- g) os *frameworks* utilizados estão propostos nos trabalhos de Carlson *et al.*(2011), Calzada Prado; Marzal (2013), Carlson; Johnston (2015), FOSTER (2015), Jeffryes; Johnston (2013) e VITAE (2011);
- h) os dados secundários serão reusados dos trabalhos de Berghmans *et al.*, (2017 [*datasets*]), de Kim (2016 [*datasets*];) e de Kim (2017 [*datasets*]);
- i) os recursos de aprendizagem *online* a serem analisados são os ofertados nos projetos DataONE, no Projeto Data Information Literacy (DIL Project), no projeto New England Collaborative Data Management Curriculum (NECDMC), no projeto MANTRA, além de outros descritos no Quadro 4;
- j) o Público Alvo como universitários, estudantes de pós-graduação e pesquisadores, bibliotecários e demais profissionais da informação, em consonância com os recursos de aprendizagem online do Quadro 4; o principal marco teórico da perspectiva de letramento

em dados está fundamentado nas melhores práticas da ACRL (ASSOCIATION OF COLLEGE AND RESEARCH LIBRARIES, 2000; 2012; 2015).

O resumo dos conceitos que sustentam essa pesquisa são apresentados no Quadro 5, como direcionadores para a proposta conceitual da Tese.

Quadro 5 - Resumo do alinhamento conceitual

(continua)

Teorias De Base	Principais Conceitos	Referências
Acesso Aberto	Benefícios como maior visibilidade, a melhora no processo de pesquisa e a redução de custos	(DAVIS <i>et al.</i> , 2008).
Ciência Aberta	Ciência aberta sob uma perspectiva de cidadania e de democracia pelas escolas democrática e a pragmática, que consideram os dados abertos;	(FECHER; FRIESIKE, 2014).
Comunicação Científica	Comunicação Científica como rede	(ROSENDAAL; GEURTS, 1997); (BORGMAN, 2000; 2007).
	Elementos da rede de comunicação Científica	(SCHONFELD, 2013; PROSSER, 2013; BORGMAN, 2007; KILEY, 2013, CARTER, 2013; MCGRATH, 2013; BROWN, 2013).
Dados de pesquisa – ciclo de vida	Ciclo de vida da gestão de dados que ser incorporado à própria metodologia da pesquisa	(INTER-UNIVERSITY CONSORTIUM FOR POLITICAL AND SOCIAL RESEARCH, 2012).
Dados de pesquisa - conceito	Conceituação está vinculada à finalidade e não à forma	(THE ENGINEERING AND PHYSICAL SCIENCES RESEARCH COUNCIL, 2018, <i>online</i>);
Dados de pesquisa - conceito - Reúso	Reúso nas Ciências Sociais	(BORGMAN, 2000; 2007); (TENOPIR, 2011, 2015); (CURTY, 2016a, CURTY, 2016b, CURTY <i>et al.</i> , 2017).
Frameworks em letramento de dados	Quadros conceituais de letramento com características gerais	(CARLSON <i>et al.</i> , 2011; CALZADA PRADO; MARZAL, 2013; JEFFRYES; JOHNSTON, 2013; CARLSON; JOHNSTON, 2015; VITAE, 2011; FOSTER, 2015),
Letramento em dados	Conceitos e práticas	(CARLSON; JOHNSTON, 2015); (CARLSON <i>et al.</i> , 2011); (KOLTAY, 2015a; 2016b; 2017a; 2017b); (SCHNEIDER, 2013);
Público Alvo	Universitários, estudantes de pós-graduação e pesquisadores, bibliotecários e demais profissionais da informação.	Projetos DataONE, Data Information Literacy (DIL Project), New England Collaborative Data Management Curriculum (NECDMC), MANTRA e demais do Quadro 4.
Metodologia de Pesquisa	Pesquisadores e bibliotecários	(BARDIN, 1995).
	Fundamentos e técnicas de pesquisa	(MARCONI; LAKATOS, 2007; 2012; 2017); (GIL, 2010).

(conclusão)

Teorias De Base	Principais Conceitos	Referências
Recursos de Aprendizagem online	Ambientes Virtuais de Aprendizagem	(UNIVERSITY OF OXFORD, 2018-2019); (CONSORTIUM OF EUROPEAN SOCIAL SCIENCE DATA ARCHIVES, 2018), (JOHNSTON; JEFFRYES, 2012); (DATAONE, 2012); (DIGITAL CURATION CENTER, 2014-2019); (RESEARCH DATA NETHERLANDS, <i>online</i>); (FOSTER, 2015), (CORTI <i>et al.</i> , 2014). (UNIVERSITY OF EDINBURGH, 2017); (NECDMC, 2012); (CURATING RESEARCH ASSETS AND DATA USING LIFECYCLE EDUCATION (CRADLE); UNIVERSITY OF NORTH CAROLINA; UNIVERSITY OF EDINBURGH, 2017). Quadro 4.
Teoria Ator-Rede	Agência não humanos	(LATOUR, 1993, 1995, 2001); (CALLON, 1986, 1993).
	Caixa preta	(LATOUR, 2000); (LATOUR; CALLON, 1981); (CALLON; LAW; RIP, 1986).
	Dinâmica da rede – rede de atores	(LAW, 1993); (LATOUR, 1986; 2006).
	Elementos heterogêneos - interações	(LAW, 1993); (CALLON, 1993).
	Simetria generalizada	(LATOUR, 1992); (CALLON, 1986).
	Tradução	(CALLON, 1986).

Fonte: Autoria própria (2018).

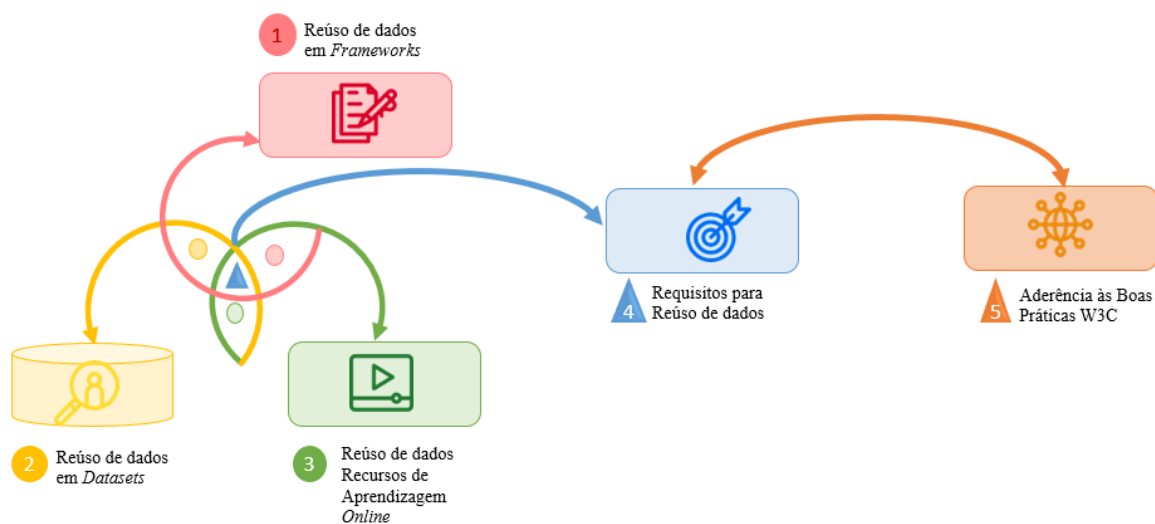
Assente nos elementos direcionadores, a proposta conceitual desta pesquisa é composta por cinco etapas. As três primeiras (etapas 1, 2 e 3) formam a análise em triangulação dos elementos do letramento informacional em dados contidas nos (1) *frameworks*, nos (2) dados secundários (*datasets*) (KIM, 2016[*datasets*];) (KIM, 2017[*datasets*];); (BERGHMANS *et al.*, 2017 [*datasets*]), e (3) nos recursos de aprendizagem *online*, sem quaisquer distinções entre esses atores.

A partir dessa análise, a próxima etapa (4) é identificar os requisitos necessários para a proposta conceitual de um ambiente virtual para o letramento em reuso de dados contextualizadas nas disciplinas das Ciências Sociais. As lentes da TAR permitem que esses atores sejam seguidos, suas formas de agência identificadas e suas dinâmicas e influências na rede sejam rastreadas, corroborando com os objetivos desta pesquisa.

Por último, na etapa 5, o resultado da análise será validado pela aderência às 35 recomendações de boas práticas do World Wide Web Consortium (W3C) (LÓSCIO; BURLE; CALEGARI, 2017), disponível no Apêndice R. Essas práticas incentivam e permitem a expansão contínua da Web como um meio para a troca de dados devido à crescente publicação de dados abertos e de dados de pesquisa (LÓSCIO; BURLE; CALEGARI, 2017). As cinco

etapas da proposta conceitual estão representadas na Figura 12.

Figura 12 - Proposta conceitual da pesquisa



Fonte: Autoria própria (2018).

Exposto o alinhamento conceitual, o próximo capítulo descreve a Metodologia para propor os requisitos de um ambiente virtual para o letramento informacional em dados, com as competências essenciais para o reúso de dados em Ciências Sociais.

5 METODOLOGIA DA PESQUISA

Os procedimentos propostos para permitir o alcance do objetivo deste trabalho estão descritos neste Capítulo. O processo metodológico apresentado cobre o planejamento da pesquisa, os procedimentos de levantamento e de revisão sistemática da literatura, a metodologia e as técnicas para análise dos conteúdos, registrando a execução operacional da pesquisa.

A seção 5.1 apresenta a classificação formal da pesquisa. A seção 5.2 detalha as macro etapas, como o planejamento, o desenvolvimento e os resultados da pesquisa. Os procedimentos para o levantamento bibliográfico estão na seção 5.3, tendo seus resultados descritos na seção 5.4. Os procedimentos da revisão sistemática da literatura estão na seção 5.5, seguido dos resultados na seção 5.6. A metodologia selecionada para a análise de dados qualitativos está descrita na seção 5.7, tendo os procedimentos de coleta e análise desses dados descritos na seção 5.8. Sequencialmente, são apresentados os procedimentos para as análises dos *frameworks* na seção 5.9, dos conjuntos de dados secundários na seção 5.10 e dos recursos de aprendizagem *online* na seção 5.11.

5.1 CLASSIFICAÇÃO FORMAL DA PESQUISA – SINTESE DESCRITIVA

Gil (2010) classifica uma tese quanto a diversos aspectos: à natureza e a utilização dos resultados, o propósito, o método utilizado em relação às técnicas de apreensão e classificação da procedência dos dados, o delineamento e a abordagem de análise dos mesmos.

A finalidade prática desta pesquisa – natureza e utilização dos resultados – a justifica como aplicada por tratar da resolução de um problema específico, o acesso aberto e o reúso de dados de pesquisa, fundamentando-se em autores como Borgman (2000; 2007), Curty (2016a, CURTY, 2016b; CURTY *et al.*, 2017) e Tenopir (2011, 2015). A pesquisa utiliza de elementos teóricos e empíricos sobre dados de pesquisa, evidenciando, a partir destes, requisitos para o letramento em reúso de dados nas Ciências Sociais.

Esta pesquisa pode ser considerada quanto aos propósitos como eminentemente descritiva, já que é necessário coletar e registrar de forma organizada os dados e os problemas que não estão explicitados cabalmente em documentos, antes estão presentes subliminarmente no ambiente investigado, o que justifica-se no uso de dados secundários sobre o reúso de dados. A característica dos estudos descritivos é também oferecer maior fundamento, clareza e

completude à pesquisa, com a descrição efetiva dos elementos coletados e analisados (GIL, 2010).

Para Gil (2010), a pesquisa descritiva também busca descobrir a existência de associações entre variáveis, como neste trabalho, o que se evidencia pela análise em triangulação de três variáveis: os resultados do levantamento bibliográfico por meio dos *frameworks* selecionados, as suas relações com os conjuntos de dados e com os recursos de aprendizagem *online*.

No entanto, o propósito exploratório também se justifica, pois evidencia-se, segundo Gil (2010), no estabelecimento de novas relações entre variáveis estudadas, caracterizando-as e descrevendo-as, com vistas a uma melhor compreensão dos fenômenos, bem como na busca de novos métodos e técnicas em uma área pouco explorada, neste caso a proposta conceitual de requisitos de um ambiente virtual para o letramento em reuso de dados, conforme demonstrado nos resultados do levantamento bibliográfico na seção 1.5 de forma resumida e mais detalhada na seção 5.3 deste Capítulo.

Os estudos exploratórios visam ainda clarificar e aprimorar ideias e conceitos, proporcionando “familiaridade com o problema em questão” (GIL, 2010, p. 27), bem como com o ambiente (MARCONI; LAKATOS, 2012, p. 71), mas aprofundando a abordagem ao trazer novos olhares, novos métodos ou soluções inovadoras para as questões estudadas. Logo, este estudo também se justifica como exploratório por examinar uma abordagem pouco comum, tendo em vista que está contextualizadas na disciplina das Ciências Sociais e sob a perspectiva da Teoria Ator-Rede, conforme também demonstrado no levantamento bibliométrico – Seção 1.5 do Capítulo 1.

Pelas características flexíveis dos estudos exploratórios que, segundo Gil (2010), incluem variados aspectos relativos ao fenômeno estudado, como o levantamento bibliográfico, e/ou o relato de pessoas com experiências práticas e as análises de exemplos que estimulam a compreensão do objeto de estudo esta pesquisa, ainda que eminentemente descritiva contém, portanto, elementos da pesquisa exploratória. A variedade de conteúdos presentes sobre o letramento em dados, como os *frameworks* identificados no levantamento bibliográfico, a experiência prática dos pesquisadores expressa nas respostas dos questionários reutilizados e os requisitos disponíveis nos recursos de aprendizagem *online*, são exemplos de elementos exploratórios que aproximam a compreensão e promovem a familiaridade com o problema em questão.

Diante do exposto, a tipologia do presente estudo pode ser considerada exploratório-descritivo combinada, contudo predominantemente descritiva como já pontuado, em que se

descreve um fenômeno por análises empíricas – conjunto de dados secundários – e teóricas – *frameworks* (MARCONI; LAKATOS, 2012). As análises empíricas apoiaram-se nos conjuntos de dados – *datasets* para reúso, as análises teóricas nos estudos que elaboraram os *frameworks* e nos recursos de aprendizagem *online*.

Quanto ao Métodos e aos procedimentos de coleta de dados, realizou-se um estudo bibliográfico, utilizando fontes secundárias, como artigos científicos, livros, teses, relatórios técnicos e outros materiais bibliográficos por meio da técnica de Revisão Sistemática da Literatura (RSL), selecionando o conteúdo teórico pertinente ao tema da pesquisa. Para as técnicas de apreensão de dados utilizou-se da recuperação de dados secundários de pesquisa relacionados com o objeto deste estudo.

Quanto à classificação da procedência dos dados, de acordo com Marconi e Lakatos (2012), estes são classificados como dados secundários. Para Lakatos e Marconi (2012), uma das características dos dados secundários é a extensão do campo de estudos dos mesmos. Os dados secundários usados nesta pesquisa são resultados de levantamentos e de *surveys*, obtidos pela consulta direta aos pesquisadores, cujo comportamento se desejava conhecer. Os conjuntos de dados de pesquisa secundários utilizados como os de Kim (2016 [*datasets*]), os de Kim, (2017 [*datasets*]) e os de Berghmans *et al.* (2017 [*datasets*]) foram selecionados por meio de acesso, e da disponibilidade, no Repositório de dados de pesquisa em Ciências Sociais Inter-university Consortium for Political and Social Research (ICPSR). Segundo a classificação de Lakatos e Markoni (2012), a natureza dos dados secundários pode ser subjetiva ou de opiniões e atitudes, que é o caso do presente estudo, pois os levantamentos e os *surveys* reutilizados referem-se aos estudos comportamentais sobre as percepções dos pesquisadores que estão familiarizados com o acesso e o reúso de dados de pesquisa.

Quanto aos critérios de seleção dos conjuntos de dados secundários, esta se deu de forma não probabilística, sendo por tipicidade. A amostragem não probabilística tem como característica a não aplicação de fórmulas estatísticas para o seu cálculo, não podendo ser objeto de tratamento estatístico em relação ao universo, como por exemplo a impossibilidade de cálculo de erros de amostragem e desvios padrão (MARCONI; LAKATOS, 2012).

A escolha por tipicidade se justifica pelo conjunto de propriedades de uma comunidade que se comporta de forma típica em relação a determinadas características, inferindo-se, portanto, que irá se comportar da mesma forma em relação ao objeto estudado, permitindo a generalização dos resultados (MARCONI; LAKATOS, 2012). A escolha por tipicidade dos conjuntos de dados levou em consideração pesquisas realizadas com pesquisadores familiarizados com a prática de depósito e/ou reúso de dados. O grau de generalização dos

resultados se deu a partir dos dados por amostragem não probabilística ou aleatória (MARCONI; LAKATOS, 2012). A seleção por tipicidade se dá na impossibilidade de seleção de uma amostra probabilística e quando possíveis erros não apresentam maior gravidade à população, ou ao entendimento e à generalização de resultados.

Quanto a seleção da amostra dos dados secundários, é importante salientar que estas foram estabelecidas pelos respectivos autores dos trabalhos. No relatório de pesquisa da Leiden University e Elsevier (2017) cujos dados de pesquisa estão publicados como Berghmans *et al.* (2017³⁰), não é possível conhecer, na sua totalidade, qual é a população global de pesquisadores brasileiros, isso porque o relatório usa como fonte os dados da UNESCO em que não consta o número de pesquisadores para o Brasil (UNITED NATIONS EDUCATIONAL, SCIENTIFIC AND CULTURAL ORGANISATION, 2015).

Contudo, a amostra do relatório da Leiden University e Elsevier (2017) é calculada sobre a população global de pesquisadores em todo o mundo, com 1.162 respostas, representando uma taxa de resposta de 2,3% (LEIDEN UNIVERSITY; ELSEVIER, 2017). Mesmo que essa amostra não seja estatisticamente representativa da população de análise, a validade é admitida dentro de um contexto específico (MARCONI; LAKATOS, 2012). Salienta-se que as amostras utilizadas nos dados secundários dos trabalhos de Kim (2016³¹) e Kim (2017³²) foram validadas nos originais.

Outra etapa realizada como procedimento para a coleta de dados foi a busca e a localização de dados de pesquisa sobre o objeto de estudo. Ainda que os dados não tenham sido levantados por esta pesquisadora de forma direta (pesquisa de campo como questionários e/ou entrevistas), havia a disponibilidade de dados levantados pela consulta direta às pessoas cujo

³⁰ A pesquisa da Leiden University com a Elsevier realizada *online* foi realizada em junho-julho de 2016 com pesquisadores em todo o mundo e em todos os campos científicos, representando uma taxa de resposta de 2,3% dos 7.8 milhões de pesquisadores em tempo integral no mundo em 2013 segundo a UNESCO (UNITED NATIONS EDUCATIONAL, SCIENTIFIC AND CULTURAL ORGANISATION, 2015, p. 32). Para este estudo, as respostas são ponderadas para ser representativa da população pesquisada (LEIDEN UNIVERSITY; ELSEVIER, 2017, p. 20).

³¹ A população alvo da pesquisa de Kim (2016) incluiu membros do corpo docente e pesquisadores de pós-doutorado em instituições acadêmicas norte-americanas, cuja amostragem foi identificada a partir da lista de estudiosos do Banco de Dados do Acadêmico da Comunidade de Ciências (CoS) (<http://pivot.cos.com>), que fornece um diretório de perfis de pesquisadores no mundo principalmente de universidades e faculdades. O instrumento final de pesquisa de campo foi distribuído aos 16.165 participantes potenciais da pesquisa em 56 disciplinas. Foram recebidas 2.470 respostas válidas para a análise inicial dos dados (15,28% da taxa de resposta) (KIM, 2015).

³² Foram considerados como população-alvo para a pesquisa de Kim (2017) os cientistas sociais nos EUA, extraídos do Banco de dados Science / Scholars (COS), que inclui membros do corpo docente, pesquisadores de pós-doutorado e pesquisadores de pós-graduação utilizado para o seu quadro de amostragem. Com base na amostragem, um total de 2.285 acadêmicos foram aleatoriamente selecionados para a amostra da pesquisa dentre uma população de 77.784 pesquisadores no banco de dados do COS (em 16 de setembro de 2012). Um total de 1.528 respostas válidas foram recebidas, sendo utilizadas 1.237 para a análise (KIM, 2017, p. 5).

comportamento se desejava conhecer, nos repositórios de dados de pesquisa, caracterizando efetivo reúso de dados, delimitação primeva deste estudo.

Prosseguindo, quanto a abordagem de análise dos dados, o estudo é qualitativo, utilizando a técnica de análise de conteúdo de Bardin (1995), detalhada na seção 5.7, tanto para a análise dos *frameworks* quanto para os conjuntos de dados de pesquisa secundários. Bardin (1995) sustenta que esta análise consiste em transformar grandes quantidades de dados brutos em informações concisas a respeito de situações, de acontecimentos e de pessoas, que expressam significações e possibilitam a interpretação. Nessa técnica são propostas práticas que minimizam as incertezas geradas a partir da análise subjetiva, buscando evidenciar as inter-relações evidentes.

O processo de estudo dos dados se deu de forma comparativa entre os *frameworks*, os dados secundários e os recursos de aprendizagem *online* pelo uso da técnica de análise conteúdo, igualmente na perspectiva de Bardin (1977, 2011), apoiada no uso do *software* de análises NVivo. O resumo de classificação metodológica para este trabalho é apresentado no Quadro 6.

Quadro 6 - Resumo de enquadramento metodológico

(continua)

Crítérios de Classificação	Classificação	Evidência
Natureza	Aplicada	A pesquisa é aplicada ao acesso aberto e reúso de dados de pesquisa;
Propósitos	Exploratório	-A pesquisa utiliza-se de elementos teóricos e empíricos sobre para a propor os requisitos; -Examina uma abordagem pouco explorada, contextualizadas na disciplina das Ciências Sociais e sob a perspectiva da Teoria Ator-Rede; -Características flexíveis e variadas da pesquisa exploratória evidenciadas pelos frameworks, a experiência dos pesquisadores e os recursos de aprendizagem online, que aproximam a compreensão e promovem a familiaridade com o fenômeno estudado;
	Descritivo	-Busca descobrir a existência de associações entre os frameworks e as suas relações com os conjuntos de dados e com os recursos de aprendizagem online; -Coleta e registra de forma organizada os dados e os problemas que não são identificados em documentos, como as percepções dos pesquisadores contidas nos questionários reutilizados;
	Exploratório-descritivo combinado	-É exploratório-descritivo combinada, predominantemente descritiva, pois descreve-se um fenômeno por análises empíricas e teóricas; -As análises empíricas apoiam-se nos conjuntos de dados secundários para reúso; -As análises teóricas apoiam-se nos estudos que elaboraram os frameworks e nos recursos de aprendizagem online;
Métodos e procedimentos de coleta de dados	Bibliográfico e de campo	É um estudo bibliográfico, utilizando fontes secundárias; Dados de pesquisa de campo (com reúso de dados secundários);

(conclusão)

Critérios de Classificação	Classificação	Evidência
Técnicas de apreensão	Revisão Sistemática da Literatura (RSL) e Recuperação de Dados secundários	-Seleção Bibliográfica por meio da Revisão Sistemática da Literatura (RSL); -Recuperação de dados secundários de pesquisa relacionados com o objeto deste estudo;
Procedência dos dados	Dados secundários	Usa dados secundários do Repositório de dados em Ciências Sociais Inter-university Consortium for Political and Social Research (ICPSR);
i) Extensão do campo de estudos dos dados secundários	Levantamentos e <i>surveys</i>	Usa dados secundários que estudam comportamento dos pesquisadores (KIM, 2016 [datasets]; 2017 [datasets]; BERGHMANS <i>et al.</i> , 2017 [datasets]);
ii) Natureza dos dados	Pesquisa subjetiva ou de opiniões e atitudes	Usa dados quantitativos sobre as percepções e os comportamentos dos pesquisadores que estão familiarizados com o acesso e o reuso de dados;
Seleção da amostra	Não probabilística por tipicidade (em relação aos conjuntos de dados secundários)	A tipicidade dos conjuntos de dados leva em consideração pesquisas realizadas com pesquisadores familiarizados com a prática de depósito e/ou reuso de dados.
Grau de generalização dos resultados	Amostragem não probabilística ou aleatória	A generalização dos resultados se justifica pela familiaridade dos pesquisadores entrevistados, cujo comportamento de forma típica em relação à determinadas características permite a inferência de que outros pesquisadores sob as mesmas condições irão se comportar da mesma forma em relação ao objeto estudado.
Abordagem de análise dos dados	Qualitativa	Utiliza a técnica de análise de conteúdo de Bardin (1995) detalhada na seção 5.7.
Processo de estudo	Comparativo	Análise comparativa entre os frameworks, os dados secundários e os recursos de aprendizagem online pelo uso da técnica de análise conteúdo, apoiada no uso de <i>software</i> de análises NVivo.

Fonte: Adaptado de Gil (2010) e de Lakatos e Markoni (2012; 2017).

Na seção seguinte são apresentadas as macro etapas da pesquisa, desde a revisão bibliográfica, passando pela análise da literatura, localização dos dados, análise comparativa das competências de letramento em dados de pesquisa e a proposta de requisitos para um ambiente virtual para o letramento em reuso de dados contextualizadas nas disciplinas das Ciências Sociais.

5.2 MACROETAPAS DE PESQUISA

As macro etapas desta pesquisa foram realizadas em três fases: o planejamento da pesquisa (5.2.1), o desenvolvimento (5.2.2) e a apresentação dos resultados (5.2.3). A seguir é detalhada a execução de cada uma das atividades elencadas.

5.2.1 Planejamento da pesquisa

O planejamento da pesquisa se consolidou com a elaboração do projeto da pesquisa, a definição do tema, a delimitação, a problematização, os objetivos, as justificativas e com o plano de gestão de dados (*Data Management Plan – DMP*), disponível no Apêndice D.

O projeto de pesquisa foi elaborado a partir de leituras preliminares de fontes internacionais sobre as tendências em Ciência da Informação no ensino superior, especialmente no âmbito das Bibliotecas acadêmicas, área de atuação desta pesquisadora. A partir das complexidades do contexto da comunicação científica do século XXI, consolidou-se a escolha do objeto de pesquisa.

O projeto foi construído nos moldes das recomendações de Luna (2000): definir o tema, a delimitação, a problematização, os objetivos e as justificativas, o embasamento teórico e o cronograma. Esta etapa determinou a construção do capítulo 1, sendo os principais pontos aqui resgatados: o tema refere-se ao conceito-chave do acesso aberto aos dados de pesquisa, aspecto inserido no contexto da Ciência Aberta. A delimitação está vinculada ao letramento informacional em dados para o uso e reúso dos dados no âmbito das Ciências Sociais.

A problematização se insere no atual cenário da comunicação científica, em que órgãos de fomento e revistas acadêmicas passam a exigir cada vez mais o depósito dos dados de pesquisa em acesso aberto de um lado, e de outro lado, tem-se o pesquisador que precisa de capacitação para estar em conformidade com essas novas exigências. Como consequência dessas regras, são disponibilizados por meio de ambientes virtuais de pesquisa, uma ampla gama de conjuntos de dados disponíveis para o reúso, oferecendo aos pesquisadores as vantagens dessa prática.

Elaborou-se igualmente um plano de Gestão de Dados (*Data Management Plan – DPM*). Um plano de gestão de dados (DMP) é um documento formal que descreve como lidar com dados durante e após a execução da pesquisa, sendo adicionado ao projeto de pesquisa concluído³³. É um documento exigido por diversas agências de financiamento, que deve ser entregue em conjunto com o projeto de pesquisa e está disponível no Apêndice D.

Hernández Sampieri, Fernández Collado e Baptista Lucio (2013) propõem apresentar uma matriz de consistência, que relaciona de forma sintética as dimensões da pesquisa, como o problema, a pergunta de pesquisa, os objetivos, os pressupostos, as variáveis de pesquisa, o universo da amostra, o tipo de pesquisa e o método, estabelecendo um panorama do

³³ <https://goo.gl/kRa5Go>

planejamento da pesquisa. A matriz de consistência deste trabalho é apresentada no Quadro 7.

Nos estudos qualitativos, esclarece-se, é de praxe utilizar-se o termo categoria de análise em lugar do termo variáveis. No entanto, cumpre lembrar que as variáveis qualitativas, também chamadas de categóricas, são características sem valores quantitativos, mas com categorias de classificação por tipo ou atributo, por exemplo, podendo ser nominais ou ordinais (REIS; REIS, 2001).

Quadro 7 - Matriz de consistência

Problema	Objetivos	Pressupostos	Variáveis de pesquisa	Universo e amostra	Tipo de pesquisa
	Objetivos Específicos				
Há dificuldades dos pesquisadores na gestão dos dados, principalmente questões ligadas ao reuso dos mesmos. Esse cenário é mais complexo nas ciências Sociais, cujos dados são mais heterogêneos.	<p>a) OE1 - Caracterizar ambientes virtuais de pesquisa;</p> <p>b) OE2 - Levantar, nos quadros conceituais (<i>frameworks</i>) de letramento informacional em dados, os requisitos propostos para o reuso de dados;</p> <p>c) OE3 - Identificar as dificuldades no reuso dados de pesquisa, por meio de dados secundários (<i>datasets</i>);</p> <p>d) OE4 - Identificar as principais competências para o reuso de dados em recursos de aprendizado <i>online</i>;</p> <p>e) OE5 - Comparar as características do letramento, propondo um modelo de requisitos com vistas ao reuso de dados nas Ciências Sociais;</p> <p>f) OE6-Validar os requisitos propostos por meio da verificação de aderência às melhores práticas da área recomendadas pelo World Wide Web Consortium (W3C).</p>	<p>a) Sobrecarga de informações do mundo digital e na produção científica, dificulta determinar a finalidade, a qualidade e confiabilidade dos recursos disponíveis;</p> <p>b) Pesquisadores precisam do letramento informacional em dados que lhes permitirão o acesso e no reuso dos mesmos;</p> <p>c) O reuso de dados disponíveis é prejudicado pela falta de competências em letramento informacional em dados, especialmente nas Ciências Sociais devido a heterogeneidade de dados;</p> <p>d) A biblioteca acadêmica é o locus adequado para promover o letramento da gestão, do compartilhamento e do reuso de dados de pesquisa, considerando seu papel social e de responsabilidade com a geração de conhecimento e não apenas com o depósito do mesmo;</p> <p>e) A falta de letramento em dados dos pesquisadores diminui o valor potencial de reuso dos dados.</p>	<p>a) <i>Frameworks</i> de letramento em dados de pesquisa identificados na literatura Carlson <i>et al.</i> (2011); Jeffryes e Johnston (2013); Calzada Prado e Marzal (2013); Carlson e Johnston (2015); Vitae (2011) e Foster (2015);</p> <p>b) Dados secundários (<i>datasets</i>) sobre comportamento dos pesquisadores em compartilhar os seus dados (1) (KIM, 2016[<i>dataset</i>];) sobre reusar dados de outros pesquisadores (2) (KIM, 2017[<i>dataset</i>];) e sobre a reutilização de dados abertos em sua própria pesquisa (3) (BERGHMANS <i>et al.</i>, 2017[<i>dataset</i>]; LEIDEN UNIVERSITY; ELSEVIER, 2017);</p> <p>c) Recursos Virtuais de Aprendizagem como os Projetos DataONE, Data Information Literacy (DIL Project), New England Collaborative Data Management Curriculum (NECDMC), Mantra e demais do QUADRO 4.</p>	<p>a) A amostra (1) de Kim (2016) considerou membros do corpo docente e pesquisadores de pós-doutorado em instituições acadêmicas norte-americanas, de 16.165 participantes potenciais da pesquisa foram obtidas 2.470 respostas válidas;</p> <p>b) Kim (2017) em sua amostra (2) investigou cientistas sociais nos EUA, membros do corpo docente, pesquisadores de pós-graduação, 2.285 acadêmicos foram aleatoriamente selecionados dentre 77.784 pesquisadores e 361 respostas válidas foram recebidas;</p> <p>c) A Amostra (3), da Leiden University com a Elsevier (2017) eram pesquisadores em todo o mundo e de todos os campos científicos, totalizando 1.162 respostas válidas recebidas.</p>	<p>a) Pesquisa de natureza aplicada, exploratório-descritivo combinada.</p>
Pergunta de pesquisa					Método de Pesquisa
Quais são os requisitos necessários em ambientes virtuais de aprendizagem para que os pesquisadores desenvolvam competências mínimas de letramento informacional para reuso de dados de pesquisa nas Ciências Sociais, na abordagem da Teoria Ator-Rede?					<p>a) Tipo Bibliográfico e de coleta de dados de campo (com reuso de dados secundários).</p> <p>b) Técnica de tratamento: Análise de conteúdo com apoio de ferramenta de análise NVivo.</p>
Objetivos					Indicadores de Avaliação
Objetivo Geral					<p>a) <i>Frameworks</i>: características comuns dos quadros aplicáveis ao reuso de dados;</p> <p>b) Dados secundários (<i>datasets</i>): Comportamento dos pesquisadores relacionados com falta de letramento em reuso de dados;</p> <p>c) Recursos Virtuais de Aprendizagem: competências abordadas sobre o reuso de dados.</p>
Propor, a partir de elementos da Teoria Ator-Rede, requisitos para um ambiente virtual em letramento informacional de dados, com as competências mínimas para o reuso de dados em Ciências Sociais.					

Fonte: Adaptado de Hernández Sampieri, Fernández Collado e Baptista Lucio (2013).

5.2.2 Desenvolvimento da Pesquisa

O desenvolvimento da Pesquisa consistiu nas etapas do levantamento bibliográfico, de análise sistemática, de localização dos dados e de análise comparativa de conteúdo, decompondo-se em:

- a) revisão bibliográfica das teorias de base (*corpus* estático), que é o modelo teórico que permite interpretar dados e fatos coletados, orienta os pressupostos que conduzem ou fundamentam a interpretação (MARCONI; LAKATOS, 2007). Nesta pesquisa, o *corpus* estático cobriu os seguintes temas: comunicação científica, conceituação de dados de pesquisa, conceituação de letramento informacional em dados e teoria de rede de atores, resultando nos capítulos 2, 3 e 4;
- b) delimitação da estratégia de busca para o levantamento bibliográfico. A estratégia de busca foi a primeira etapa da revisão sistemática da literatura e encontra-se descrita no capítulo 1, detalhadas posteriormente nas seções 5.3.1 e 5.3.2;
- c) realização do levantamento bibliográfico da literatura do estado da prática (*corpus* dinâmico). Os procedimentos do levantamento bibliográfico estão detalhados na seção 5.3 e os seus resultados estão descritos na seção 5.4;
- d) a partir do *corpus* teórico, revisou-se de forma sistemática a bibliografia do estado da prática em Letramento Informacional em Dados (LID). A metodologia empregada para a Revisão Sistemática da Literatura (RSL) tem suas etapas desdobradas na seção 5.5.
- e) os resultados da revisão sistemática, apresentando os documentos selecionados para a análise final com a consolidação estatística do *corpus* teórico utilizado são apresentados na seção 5.6; a partir destas análises foi possível identificar os quadros conceituais (*frameworks*) em LID. Os quadros conceituais identificados na **literatura**, por questão de ordem das seções deste relatório, estão listados na seção 3.5;
- f) análise de conteúdo dos quadros conceituais (*frameworks*): a seção 5.7 apresenta a metodologia para análise de dados qualitativos e o protocolo de execução desta pesquisa. A partir destes quadros, identificou-se as competências em letramento informacional em dados; explicitou-se também as conceituações e as práticas dos serviços relacionados aos dados de pesquisa;
- g) localização dos conjuntos de dados sobre o objeto de pesquisa para reuso e novas análises. A seção 5.8 apresenta os procedimentos de coleta de dados secundários, desde a pesquisa para a localização dos conjuntos de dados até as formas de obtenção dos mesmos;

- h) análise de conteúdo dos quadros conceituais (*frameworks*). Os procedimentos de análise de *frameworks* de letramento em dados, considerando os requisitos de reúso, estão listados na seção 5.9, identificando os requisitos para o letramento em dados;
- i) análise de conteúdo de dados para identificar as categorias de letramento em dados. Os procedimentos de análise dos conjuntos de dados de pesquisa (*datasets*) para reúso e novas análises estão na seção 5.10;
- j) análise de conteúdo dos recursos de aprendizagem online: a descrição da metodologia de análise de conteúdo dos recursos de aprendizagem online, que identificou as competências em letramento informacional em dados disponíveis em Ambientes Virtuais de Aprendizagem, está disponível na seção 5.11.

5.2.3 Resultados da Pesquisa

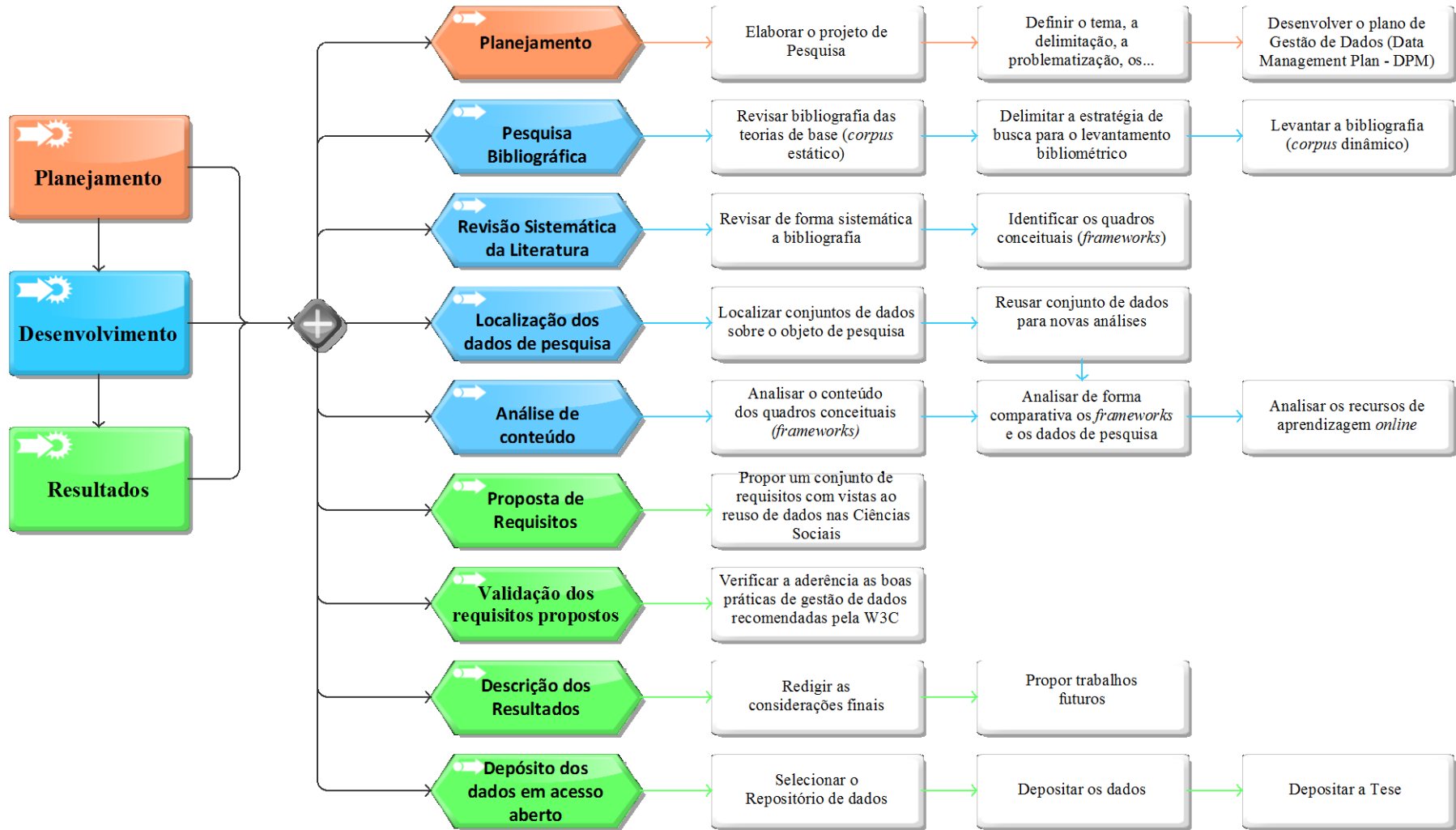
A etapa final dos resultados passa pela proposta conceitual, a descrição dos resultados e o depósito dos dados em acesso aberto:

- a) os resultados da análise qualitativa dos conjuntos de dados secundários (*datasets*); que mostrou a percepção dos pesquisadores sobre reúso de dados, estão na seção 6.1;
- b) os resultados dos recursos de aprendizagem *online* analisados, em que identificou-se as competências em letramento informacional em dados disponíveis em Ambientes Virtuais de Aprendizagem estão descritos na seção 6.2;
- c) a análise comparativa em triangulação entre os dados de pesquisa com os quadros conceituais (*frameworks*) e com os recursos *online*: fez-se a comparação dos requisitos dos *frameworks* com a percepção dos pesquisadores evidenciada em pesquisas anteriores (KIM, 2016[*datasets*];) (KIM, 2017[*datasets*];); (BERGHMANS *et al.*, 2017[*datasets*];) e com as competências identificadas nos recursos de aprendizagem *online*, procurando relações com as dificuldades identificadas no reúso dos dados, buscando relações com as competências evidenciadas pelos recursos de aprendizagem em letramento informacional de dados. Os resultados desta análise comparativa estão detalhados na seção 6.3;
- d) proposta conceitual dos requisitos do ambiente de virtual para letramento em reúso de dados. A descrição dos resultados do conjunto de requisitos propostos são detalhados na seção 6.4;

- e) a validação do conjunto de requisitos propostos por meio da verificação de aderência às melhores práticas da área, recomendadas pelo World Wide Web Consortium (W3C), estão detalhados na seção 6.5;
- f) a redação das considerações finais e das propostas de trabalhos futuros. Os resultados são consolidados estão na seção 7.1, as limitações da pesquisa foram descritas na seção 7.2 e as proposições futuras são encontradas na seção 7.3;
- g) Depósito dos dados e da Pesquisa em acesso aberto (seleção dos Repositórios de dados, depósito dos dados e da Pesquisa). Os dados de pesquisa serão depositados conforme a previsão do plano de gestão de dados (DMP), disponível no Apêndice D.

A sequência de macro etapas, etapas e atividades estão exemplificadas na Figura 13. A forma de execução dessas etapas, com o fluxo da execução das atividades da pesquisa realizadas entre 2017-2019 pode ser verificada no Apêndice E na Figura 41. Na sequência ao esquema das macro etapas, descreve-se o levantamento bibliográfico e a revisão sistemática da literatura decorrente para construção do referencial teórico.

Figura 13 - Sequência das macro etapas da pesquisa



Fonte: Autoria própria (2018).

5.3 PROCEDIMENTOS DE LEVANTAMENTO BIBLIOGRÁFICO

Para operacionalizar a revisão de literatura foi realizada um levantamento bibliográfico, que também permite demonstrar a relevância prática da pesquisa sob o aspecto da produção científica. Após os primeiros resultados parciais, foi realizada, então, a revisão sistemática da literatura (RSL) com vistas à composição do *corpus* de análise. Constituída de diversas etapas, a revisão sistemática permite a seleção dos artigos que irão compor o *corpus* teórico para a análise bibliográfica, com a leitura do texto completo de cada item para embasar a construção teórica da pesquisa (PAGANI; KOVALESKI; RESENDE, 2017), deve apresentar um método detalhado de todas as etapas e as respectivas atividades da revisão sistemática em questão (GUERTIN; BERNHARD, 2005; CANADA, 2004).

Este trabalho baseou-se nas etapas de 1 a 6 propostas na metodologia de Guertin e Bernhard (2005), que sistematizaram os conceitos do Ministério da Educação do Canadá (CANADÁ, 2004), sintetizadas na Figura 14. As etapas 1 e 2 constituem o protocolo de levantamento bibliográfico e o registro detalhado de todas as atividades de planejamento pertinentes para a elaboração da busca do *corpus* teórico da pesquisa, detalhados nesta seção. As demais etapas 3, 4 e 5, se referem às atividades de revisão sistemática dos registros recuperados, estão detalhadas na seção 5.5. A etapa 6 finaliza a análise sistemática, que consiste na redação do texto de revisão de literatura.

Figura 14 - Macro etapas de pesquisa do levantamento bibliográfico e da análise sistemática



Fonte: Adaptado de Guertin e Bernhard (2005) e Canadá (2004).

As seis macro etapas da revisão sistemática possuem desdobramentos com atividades subsequentes (GUERTIN; BERNHARD, 2005; CANADA, 2004). Os desdobramentos de cada uma das etapas e a sua forma de execução estão disponíveis em forma gráfica de fluxo de processos no Apêndice F. Essas atividades e os respectivos procedimentos operacionais estão ilustrados na Figura 15 e as atividades detalhadas nas seções subsequentes.

Figura 15 - Desdobramentos do procedimento de revisão sistemática

	Etapa 1	Etapa 2	Etapa 3	Etapa 4	Etapa 5	Etapa 6
Etapas	Etapa 1 - Especificar o assunto	Etapa 2 - Definir as estratégias de localização da informação	Etapa 3 - Selecionar os documentos	Etapa 4 - Extrair as informações	Etapa 5 - Tratar a informação	Etapa 6 - Produzir o trabalho
Atividades	<p>1.1) entender a natureza, os objetivos e a extensão do trabalho;</p> <p>1.2) tornar a questão precisa</p> <p>1.3) procurar as ideias centrais e as palavras-chave;</p> <p>1.4) fazer um plano provisório;</p> <p>1.5) planejar os recursos a serem utilizados;</p> <p>1.6) fazer o plano de trabalho de pesquisa</p>	<p>2.1) escolher as bases nas quais quer trabalhar;</p> <p>2.2) Construir as expressões da pesquisa a partir das palavras-chave;</p> <p>2.3) executar a busca nas bases de dados</p> <p>2.4) julgar a eficácia da sua pesquisa e ajustar se necessário;</p> <p>2.5) selecionar o que encontrar de importante sobre o assunto da pesquisa;</p> <p>2.6) trocar ideias com os colegas e professores;</p>	<p>3.1) organizar os documentos importantes;</p> <p>3.2) avaliar a qualidade da informação encontrada;</p> <p>3.3) identificar os elementos de informação que são pertinentes;</p> <p>3.4) registrar as referências, na íntegra, dos documentos;</p>	<p>4.1) ler atentamente;</p> <p>4.2) indicar as fontes encontradas;</p> <p>4.3) organizar as notas;</p> <p>4.4) fazer avaliação se esta no caminho certo.</p>	<p>5.1) comparar e avaliar as informações obtidas a partir de fontes diferentes;</p> <p>5.2) analisar o que encontrou esta dentro do assunto procurado;</p> <p>5.3) sintetizar as ideias e as informações provenientes de fontes diferentes.</p>	<p>6.1) sintetizar as informações disponíveis e realizar as inferências que possam contribuir na elaboração da pesquisa;</p> <p>6.2) escrever o referencial teórico.</p>

Fonte: Adaptado de Guertin e Bernhard (2005) e Canadá (2004).

5.3.1 Especificação do assunto - Etapa 1

A partir de leituras da literatura especializada, identificou-se que temas como Ciência Aberta e Dados de Pesquisa eram tendência no campo da Ciência da Informação. Um aprofundamento possibilitou identificar que havia necessidade de maiores estudos em relação ao letramento para essas novas abordagens da ciência no século XXI, em razão da mudança do paradigma científico na geração e no uso intensivo de dados, ou a *eScience*.

Como o tema estava circundado de complexidades e envolvia muitos aspectos interdependentes, selecionou-se a Teoria Ator-Rede como viés metodológico para amparar a fundamentação. A atribuição da delimitação da pesquisa contribuiu para a precisão do objeto de estudo.

A investigação preliminar deste trabalho se deu a partir de buscas isoladas para cada um dos cinco grandes conceitos-chaves relacionados com o objeto de pesquisa: (i) ciência aberta; ii) dados abertos de pesquisa; (iii) letramento informacional de dados; (iv) teoria ator-rede e (v) ambiente virtual de pesquisa.

A partir do levantamento preliminar, foram selecionados os conceitos-chave e os descritores que seriam utilizados no levantamento bibliográfico. Os termos de busca escolhidos foram os mais recorrentes na literatura internacional, como: (a) *open science*, (b) *research data* e *scientific data*, (c) *data literacy* e *data information literacy*, (d) *Actor-network theory* e (e) *virtual research environment*.

O planejamento do levantamento bibliográfico foi constituído em função dos tipos de documentos a serem localizados, sendo na primeira fase executada a busca de teses e de dissertações e na segunda fase foi realizada a busca de artigos e de *e-books*.

Nos recursos a serem utilizados foram privilegiados os de acesso aberto ou com acesso restrito no âmbito das instituições de ensino que a pesquisadora possui vínculo, o que viabilizaria o acesso aos recursos informacionais de pesquisa. Sequencialmente definiu-se estratégias de busca e localização de informações.

5.3.2 Definição das estratégias de busca e recuperação da informação - Etapa 2

As estratégias de busca se constituem pela seleção das fontes de informação para a pesquisa e pela construção das expressões de busca que traduzam a forma como os conceitos se relacionam, sob a perspectiva do que se quer investigar na pesquisa (GUERTIN; BERNHARD, 2005; CANADA, 2004).

As bases pesquisadas foram selecionadas nos seguintes portais de pesquisa: (i) O Portal de Periódicos da Capes, (ii) o Portal EBSCOHost e a (iii) Base referencial Web of Science.

O Portal de Periódicos da Capes conta com 533 bases de dados, entre as de acesso aberto e restrito, mantido pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes), fundação ligada ao Ministério da Educação (MEC). O Portal é uma biblioteca virtual que reúne e disponibiliza às IES no Brasil fontes de informação científica e tecnológica internacional (BRASIL, 2015, *online*).

Dentre as bases de dados disponíveis no portal EBSCOHost, algumas foram selecionadas³⁴ para esse trabalho, sendo que as mais relevantes para o tema desta pesquisa foram a Information Science & Technology Abstracts (ISTA) e a Library, Information Science & Technology Abstracts with Full Text (LISTA), por cobrirem a Ciência da Informação.

A Web of Science, base multidisciplinar referencial de publicações científicas, indexa somente os periódicos mais citados em suas respectivas áreas (WEB OF SCIENCE, 2017). A estratégia de pesquisa foi reproduzida na interface de busca de cada uma das bases de dados indicadas, utilizando os mesmos campos alvos para recuperação da informação. A expressão da estratégia de busca foi elaborada para os conceitos-chave de forma isolada (busca simples) e os mesmos conceitos de forma coordenada (busca combinada). A expressão de busca, considerou, além da seleção dos conceitos e a atribuição de descritores, o estabelecimento de sinonímia a fim de não perder resultados que estivessem indexados com outros termos.

O procedimento da busca simples foi realizado da seguinte forma: para cada conceito-chave foram selecionados termos descritores e seus sinônimos. Para a realização da **busca dos termos de forma isolada, ou a busca simples**, foram elaboradas as seguintes expressões de pesquisa listadas a seguir, considerando que para cada conceito-chave da pesquisa utilizaram-se descritores para o idioma português e para o inglês:

- a) para o conceito-chave de Ciência Aberta usou-se o descritor em inglês "*Open Science*" e em português utilizou-se o descritor "Ciência Aberta";

³⁴ Bases selecionadas para esta pesquisa: Academic Search Ultimate, Academic Search Premier, Business Source Ultimate, Applied Science & Technology Source Ultimate, Computers & Applied Sciences Complete, eBook Academic Collection (EBSCOhost), eBook Business Collection (EBSCOhost), eBook Collection (EBSCOhost), Library, Information Science & Technology Abstracts with Full Text (LISTA), Education Source, E-Journals, Humanities Source Ultimate, Information Science & Technology Abstracts (ISTA), SocINDEX with Full Text, Sociology Source Ultimate. Na Atualização do levantamento em Agosto de 2019, as seguintes bases de dados não estavam mais disponíveis: Applied Science & Technology Source Ultimate, eBook Academic Collection (EBSCOhost), eBook Business Collection (EBSCOhost), Education Source e Sociology Source Ultimate.

- b) para o conceito-chave de Dados abertos de Pesquisa utilizaram-se os descritores em inglês *"open data" OR "research data" OR "scientific data"* e em português utilizaram-se os descritores "dados abertos" OU "dados de Pesquisa" OU "dados científicos";
- c) para o conceito-chave de Letramento informacional em dados os descritores em inglês *"Data literacy" OR "information data literacy" OR "data information literacy"* e em português utilizaram-se os descritores "letramento de dados" OU "letramento em dados" OU "letramento informacional de dados" OU "letramento informacional em dados";
- d) para o conceito-chave de Teoria Ator-rede utilizaram-se os descritores em inglês *"Actor-Network theory" OR "Actor Network theory"* e em português utilizaram-se os descritores "Teoria Ator-Rede" OU "Teoria Ator Rede";
- e) para o conceito-chave de Ambiente Virtual de Pesquisa utilizaram-se os descritores: *"Virtual Research Environment" OR "Research Environment"* e em Português utilizaram-se os descritores "Ambiente Virtual de Pesquisa" OU "Ambiente Virtual".

Para a pesquisa de cada conceito-chave, em inglês e português, foi utilizado o conjunto de descritores e de sinônimos apresentados no Quadro 8.

Quadro 8 - Conceitos e estratégia de busca simples

Idiomas	Conceitos-Chave	Palavras-Chave / Key-Words	Descritores/ Sinônimos
Inglês	<i>Open Science</i>	<i>Open Science</i>	<i>"Open Science"</i>
Português	Ciência Aberta	Ciência Aberta	"Ciência Aberta"
Inglês	<i>Open Research Data</i>	<i>Open Research Data</i>	<i>"Open data" OR "research data" OR "scientific data"</i>
Português	Dados abertos de pesquisa	Dados abertos de pesquisa	"Dados abertos" OR "dados de pesquisa" OR "dados científicos"
Inglês	<i>Data information literacy</i>	<i>Data information literacy</i>	<i>"Data literacy" OR "information data literacy" OR "data information literacy"</i>
Português	Dados abertos de pesquisa	Letramento em dados	"(Letramento em dados" OR "Letramento informacional" OR "Letramento informacional em dados")
Inglês	<i>Actor-Network theory</i>	<i>Actor-Network theory</i>	<i>"Actor-Network theory" OR "Actor Network theory"</i>
Português	Teoria Ator-Rede	Teoria Ator-Rede	"Teoria Ator Rede" OR "Teoria Ator-Rede"
Inglês	<i>Virtual Research Environment</i>	<i>Virtual Research Environment</i>	<i>"Virtual Research Environment" OR "Research Environment"</i>
Português	Ambiente Virtual de Pesquisa	Dados abertos de pesquisa	"Ambiente Virtual de Pesquisa" OR "Ambiente de Pesquisa"

Fonte: Autoria própria (2018).

A busca com os descritores apresentados foi repetida em cada uma das fontes selecionadas, respectivamente: (i) O Portal de Periódicos da Capes, (ii) o Portal EBSCOHost e a (iii) Base de dados referencial *Web of Science*.

A pesquisa foi realizada com a aplicação de filtros e de acordo com os seguintes parâmetros: no Portal de Periódicos da Capes, o campo utilizado para busca no texto foi o de 'Assuntos'. No portal da EbscoHost, o campo padrão selecionado para a busca foi: 'Assunto, títulos e resumo'. Já na base Web of Science, o campo de busca utilizado foi o 'Tópico', que engloba os campos Assuntos, títulos e resumo, automaticamente. Em todas as fontes pesquisadas, foram aplicados os filtros de pesquisa por período de cobertura, limitado entre os anos de 2012 e 2018, tendo sendo posteriormente atualizado até 2019³⁵. Também se adotou o critério de periódicos revisados por pares.

Para a **busca combinada** foram construídas as seguintes expressões:

- a) para a combinação (A) utilizaram-se os descritores: *"Open Science" AND ("open data" OR "research data" OR "scientific data")*. Em português usou-se: "Ciência Aberta" AND ("dados abertos" OR "dados de pesquisa" OR "dados científicos");
- b) para a combinação (B) utilizaram-se os descritores: *"Open science" AND ("open data" OR "research data" OR "scientific data") AND ("Data literacy" OR "information data literacy" OR "data information literacy")*. Em português usou-se: "Ciência Aberta" AND ("dados abertos" OR "dados de pesquisa" OR "dados científicos") AND ("letramento de dados" OR "letramento em dados" OR "letramento informacional em dados" OR "letramento informacional de dados");
- c) para a combinação (C) utilizaram-se os descritores: *("open data" OR "research data" OR "scientific data") AND ("Data literacy" OR "information data literacy" OR "data information literacy")*; Em português usou-se: ("dados abertos" OR "dados de pesquisa" OR "dados científicos") AND ("letramento de dados" OR "letramento em dados" OR "letramento informacional em dados" OR "letramento informacional de dados");
- d) para a combinação (D) utilizaram-se os descritores: *("open data" OR "research data" OR "scientific data") AND ("Data literacy" OR "information data literacy" OR "data information literacy") AND ("Actor-Network theory" OR "Actor Network theory")*. Em português usou-se: ("dados abertos" OR "dados de pesquisa" OR "dados científicos") AND ("Letramento em dados" OR "Letramento informacional" OR "Letramento informacional em dados") AND ("Teoria Ator Rede" OR "Teoria Ator-Rede");

³⁵ A pesquisa foi realizada, inicialmente, entre 2012 e 2018. A data para a busca e a recuperação dos artigos foi atualizada para até agosto de 2019. Bases pesquisadas: 1. Portal EBSCO Host: Academic Search Ultimate, Academic Search Premier, Business Source Ultimate, Computers & Applied Sciences Complete, eBook Collection (EBSCOhost), E-Journals, Humanities Source Ultimate, Information Science & Technology Abstracts (ISTA), Library, Information Science & Technology Abstracts with Full Text, SocINDEX with Full Text. 2. PORTAL da CAPES. 3. WEB OF SCIENCE.

- e) para a combinação (E) utilizaram-se os descritores: (“*Virtual Research Environment*” OR “*Research Environment*”) AND (“*open data*” OR “*research data*” OR “*scientific data*”). Em português usou-se: (“Ambiente Virtual de Pesquisa” OR “Ambiente de Pesquisa”) AND (“dados abertos” OR “dados de pesquisa” OR “dados científicos”).

Para a pesquisa de cada combinação, foi utilizado o conjunto de descritores e sinônimos apresentados no Quadro 9.

Quadro 9 - Conceitos e estratégia de busca combinada

Busca Combinada	Idioma	Descritor 1	Descritor 2	Descritor 3
Combinação A	Inglês	Open Science	Open Research Data	
	Expressão	“ <i>Open science</i> ” AND (“ <i>open data</i> ” OR “ <i>research data</i> ” OR “ <i>scientific data</i> ”)		
	Português	Ciência Aberta	Dados abertos de pesquisa	
	Expressão	“Ciência Aberta” AND (“dados abertos” OR “dados de pesquisa” OR “dados científicos”)		
Combinação B	Inglês	Open Science	Open Research Data	Data information literacy
	Expressão	“ <i>Open science</i> ” AND (“ <i>open data</i> ” OR “ <i>research data</i> ” OR “ <i>scientific data</i> ”) AND (“ <i>Data literacy</i> ” OR “ <i>information data literacy</i> ” OR “ <i>data information literacy</i> ”)		
	Português	Ciência Aberta	Dados abertos de pesquisa	Letramento em dados
	Expressão	“Ciência Aberta” AND (“dados abertos” OR “dados de pesquisa” OR “dados científicos”) AND (“ <i>letramento de dados</i> ” OR “ <i>letramento em dados</i> ” OR “ <i>letramento informacional em dados</i> ” OR “ <i>letramento informacional de dados</i> ”)		
Combinação C	Inglês	Open Research Data	Data information literacy	
	Expressão	(“ <i>open data</i> ” OR “ <i>research data</i> ” OR “ <i>scientific data</i> ”) AND (“ <i>Data literacy</i> ” OR “ <i>information data literacy</i> ” OR “ <i>data information literacy</i> ”)		
	Português	Dados abertos de pesquisa	Letramento em dados	
	Expressão	(“ <i>dados abertos</i> ” OR “ <i>dados de pesquisa</i> ” OR “ <i>dados científicos</i> ”) AND (“ <i>letramento de dados</i> ” OR “ <i>letramento em dados</i> ” OR “ <i>letramento informacional em dados</i> ” OR “ <i>letramento informacional de dados</i> ”)		
Combinação D	Inglês	Open Research Data	Data information literacy	Actor-Network theory
	Expressão	(“ <i>open data</i> ” OR “ <i>research data</i> ” OR “ <i>scientific data</i> ”) AND (“ <i>Data literacy</i> ” OR “ <i>information data literacy</i> ” OR “ <i>data information literacy</i> ”) AND (“ <i>Actor-Network theory</i> ” OR “ <i>Actor Network theory</i> ”)		
	Português	Dados abertos de pesquisa	Letramento em dados	Teoria Ator-Rede
	Expressão	(“ <i>dados abertos</i> ” OR “ <i>dados de pesquisa</i> ” OR “ <i>dados científicos</i> ”) AND (“ <i>Letramento em dados</i> ” OR “ <i>Letramento informacional</i> ” OR “ <i>Letramento informacional em dados</i> ”) AND (“ <i>Teoria Ator Rede</i> ” OR “ <i>Teoria Ator-Rede</i> ”)		
Combinação E	Inglês	Virtual Research Environment	Open Research Data	
	Expressão	(“ <i>Virtual Research Environment</i> ” OR “ <i>Research Environment</i> ”) AND (“ <i>open data</i> ” OR “ <i>research data</i> ” OR “ <i>scientific data</i> ”)		
	Português	Ambiente Virtual de Pesquisa	Dados abertos de pesquisa	
	Expressão	(“Ambiente Virtual de Pesquisa” OR “Ambiente de Pesquisa”) AND (“dados abertos” OR “dados de pesquisa” OR “dados científicos”)		

Fonte: Autoria própria (2018).

Complementando, realizou-se a pesquisa para o levantamento das teses e dissertações da área, observando-se os mesmos critérios de busca dos artigos, porém, as fontes de busca foram as bases de repositórios institucionais nacionais como o Portal OasisBR, a Biblioteca Digital Brasileira de Teses e Dissertações (BDTD), ambos do Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict) e o catálogo de Teses da Capes (Plataforma Sucupira).

Nas fontes pesquisadas (OasisBR, a BDTD e o Catálogo de Teses da Capes) foram utilizadas as estratégias de busca A, B, C, D e E, com o critério ‘Palavra-chave’ em todas as caixas de pesquisas. A abrangência da pesquisa utilizou o recorte temporal entre os anos de 2012 a 2017, e, posteriormente, foi atualizado para 2019. As teses e dissertações foram analisadas de acordo com o título a fim de se verificar a aderência do trabalho com os termos pesquisados. Quando o título não foi suficiente para entender o contexto da publicação, uma leitura detalhada do resumo foi realizada. Os resultados dessas estratégias de busca são apresentados a seguir.

5.4 RESULTADOS DO LEVANTAMENTO BIBLIOGRÁFICO

A busca no levantamento bibliográfico foi executada respeitando os parâmetros de cada fonte de informação para a que a pesquisa pudesse recuperar efetivamente os documentos cujos descritores fossem localizados em partes específicas do texto selecionado. No Portal de Periódicos da Capes, no portal da EbscoHost e na base Web of Science o parâmetro utilizado foi ‘Palavra-chave’ ou ‘Assunto’. Para todas as fontes foram selecionados os filtros de pesquisa por período de cobertura, limitado aos anos de 2012 a 2019³⁶, conforme já explicitado, bem como o critério de periódicos revisados por pares. Os números de recorrências dos conceitos-chaves pesquisados na busca simples, somados os resultados das buscas nos idiomas português e inglês são apresentados na Tabela 4.

³⁶ A pesquisa foi realizada, inicialmente, entre 2012 e 2018. A data para a busca e a recuperação dos artigos foi atualizada para até agosto de 2019. Bases pesquisadas: 1. Portal EBSCO Host: Academic Search Ultimate, Academic Search Premier, Business Source Ultimate, Computers & Applied Sciences Complete, eBook Collection (EBSCOhost), E-Journals, Humanities Source Ultimate, Information Science & Technology Abstracts (ISTA), Library, Information Science & Technology Abstracts with Full Text, SocINDEX with Full Text. 2. Portal Da Capes. 3. Web of Science.

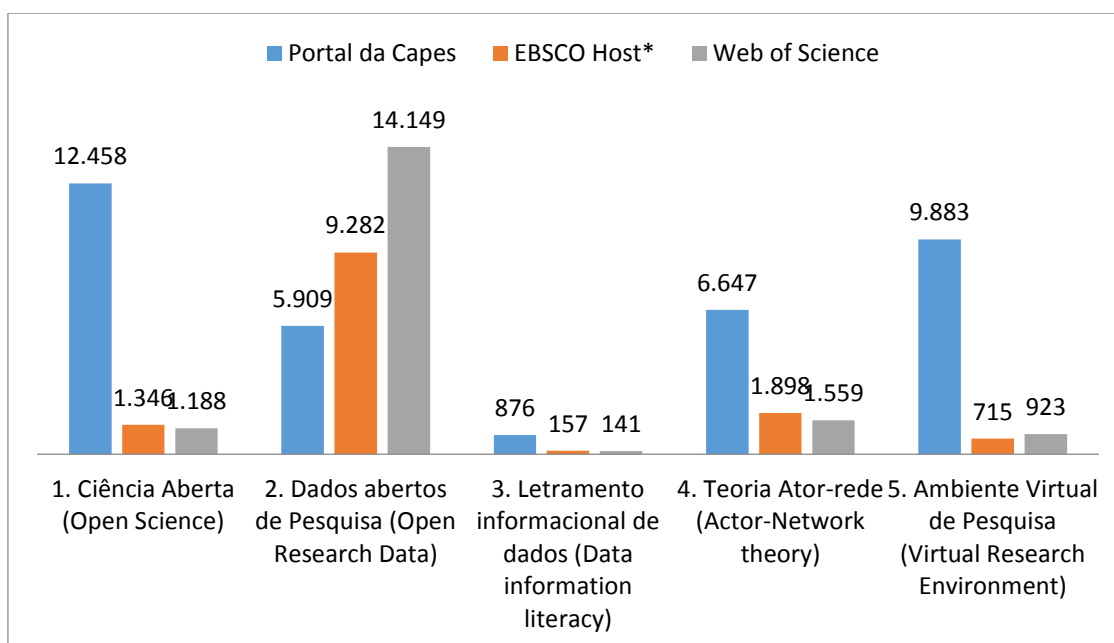
Tabela 4 - Frequência de conceitos-chaves – busca simples

Conceitos-chave 2012-2019	Bases de dados Bibliográficas		
	Portal da Capes	EBSCO Host	Web of Science
1. Ciência Aberta (<i>Open Science</i>)	12.458	1.346	1.188
2. Dados abertos de Pesquisa (<i>Open Research Data</i>)	5.909	9.282	14.149
3. Letramento informacional de dados (<i>Data information literacy</i>)	876	157	141
4. Teoria Ator-rede (<i>Actor-Network theory</i>)	6.647	1.898	1.559
5. Ambiente Virtual de Pesquisa (<i>Virtual Research Environment</i>)	9.883	715	923

Fonte: Autoria própria (2019).

A produção científica dos conceitos-chave cujos números de ocorrências foram apresentados na Tabela 4 está demonstrada no Gráfico 1. Nota-se que o conceito-chave principal desta pesquisa expresso pelos descritores *Open Research Data* é amplamente discutido na literatura quando pesquisado sem a combinação com outros conceitos.

Gráfico 1 - Levantamento bibliométrico – busca simples



Fonte: Autoria própria (2019).

Quando pesquisados de forma combinada entre si e com os outros conceitos norteadores da pesquisa, a quantidade recuperada diminui sensivelmente, como mostra a Tabela 5,

especialmente em determinadas combinações. Salienta-se que estes dados já foram apresentados no Capítulo 1, Seção 1.5.1, quando da apresentação da Justificativa Teórica relacionada à originalidade desta pesquisa. Retoma-se, neste ponto, apesar da redundância, com a função dos aprofundamentos necessários à questão metodológica.

Tabela 5 - Frequência de conceitos-chaves – busca combinada

Busca Combinada	Referências Capes	Referências EBSCOHost	Referências Web of Science	Total Referências
Combinação A	163	143	202	508
Combinação B	2	1	1	4
Combinação C	19	116 ³⁷	38	173
Combinação D	6	0	0	6
Combinação E	4	24	36	64
Total referências	194	284	277	755
% de referências por bases	27,82%	32,67%	39,51%	100,00%

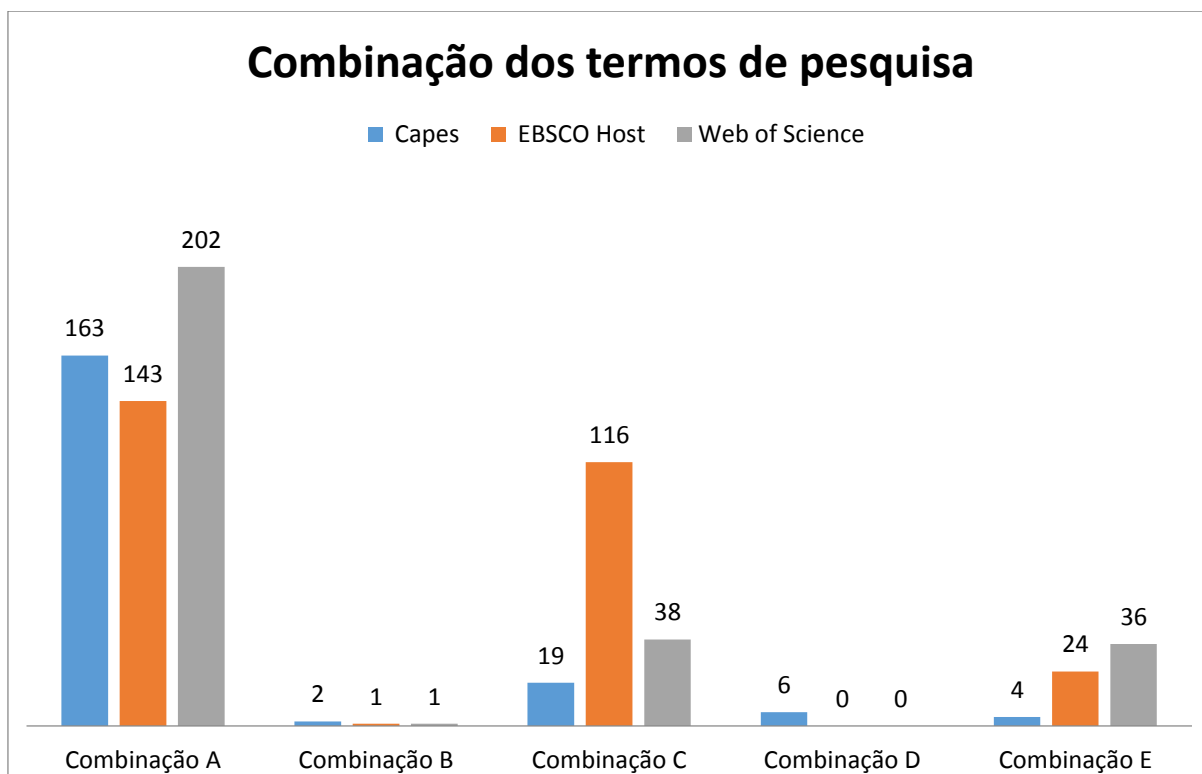
Fonte: Autoria própria (2019).

Destaca-se novamente, portanto, que as combinações B e D apresentaram muito poucos resultados. A combinação B é importante porque demonstra que não há muitos trabalhos quando se relaciona a Ciência Aberta e os Dados de Pesquisa em conjunto com o Letramento Informacional em Dados. Ou seja, ainda que haja 508 referências sobre Ciência Aberta e Dados de Pesquisa (combinação A), ao adicionar-se o conceito de Letramento Informacional em Dados (combinação B) o resultado tem uma queda expressiva, sendo de apenas quatro referências (combinação B).

A combinação D demonstra que o conceito de Dados Abertos em conjunto com Letramento Informacional em Dados, se combinados com a Teoria Ator-Rede, também tem um resultado mínimo. Comparando-se com a combinação C, esta recuperou 173 referências quando não foi incluído o termo Teoria Ator-Rede. Isso demonstra que, dentre os 173 trabalhos que tratam de letramento em dados de pesquisa, apenas 6 mencionam a Teoria Ator-Rede (resultado da combinação D). Esses resultados evidenciam uma lacuna de pesquisa e a pertinência deste trabalho, já que o mesmo aborda tanto o conceito de Letramento Informacional em Dados quanto o conceito da Teoria Ator-Rede imbricados aos temas de Ciência Aberta e de Dados de Pesquisa. A representação da combinação dos assuntos tratados neste estudo, nas bases de dados pesquisadas, é apresentada no Gráfico 2.

³⁷ Número foi mantido pois houve queda nos resultados na atualização da busca em 2019, tendo sido recuperados apenas 95 resultados, em função do cancelamento de assinatura de bases de dados.

Gráfico 2 - Descritores e combinação dos termos de pesquisa



Fonte: Autoria própria (2019).

Representa-se a composição do levantamento bibliométrico em relação aos termos mais relevantes presentes nos artigos, em uma perspectiva de evolução cronológica, conforme a legenda da evolução dos anos e cores – Figura 16. Algumas combinações foram mescladas e processadas no sistema Vosviewer³⁸ que arranja as referências de forma a mostrar as relações entre os termos mais recorrentes e o período cronológico dessas relações. Foram representadas apenas as combinações que apresentaram resultados de busca na base de dados Web of Science, agrupadas e processadas no sistema, cuja análise de resultado retornou 277 referências – Figura 16:

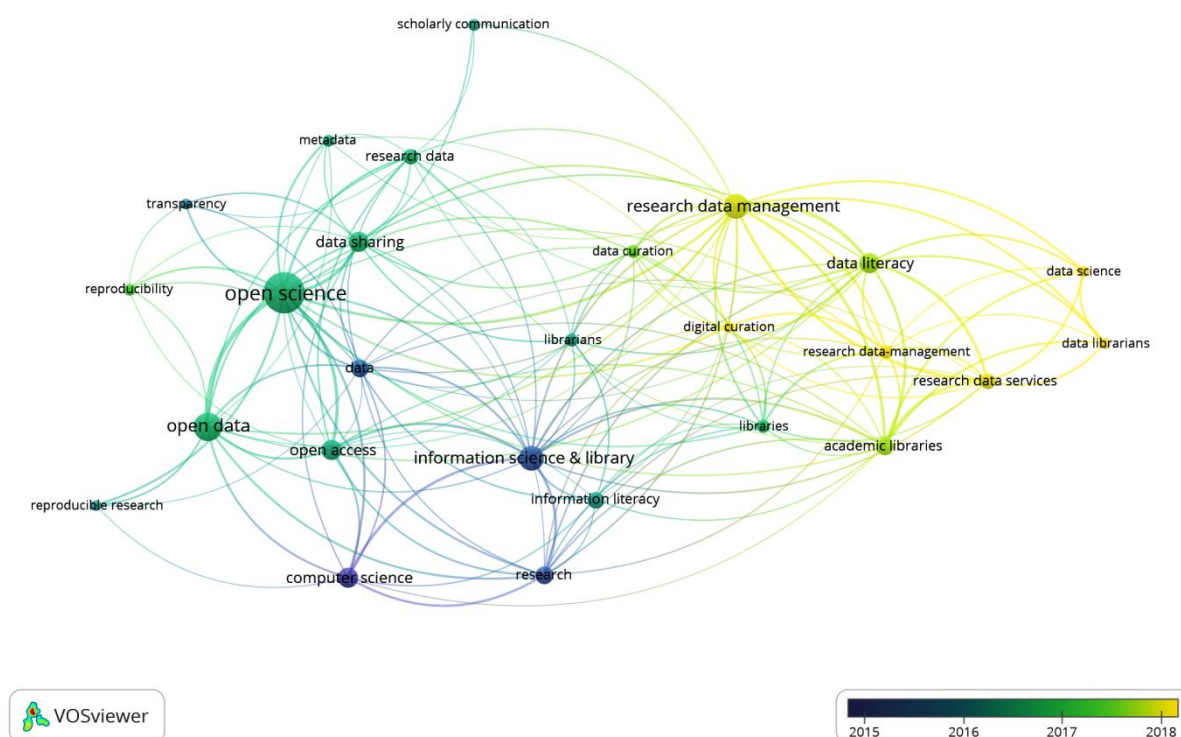
- combinação (A): *Open Science* + *Open Research Data* = 202 referências;
- combinação (B): *Open Science* + *Open Research Data* = 1 referência;
- combinação (C): *Open Research Data* + *Data information literacy* = 38 referências;
- combinação (E): *Virtual Research Environment* + *Open Research Data* = 36 referências.

Nota-se que conforme o passar do tempo existe uma aproximação diferenciada dos termos pesquisados, de acordo com um conjunto de referências analisadas pelo *software* Vosviewer, apresentados na Figura 16. Por exemplo, ao analisar-se o conjunto de referências

³⁸ <http://www.vosviewer.com>

recuperadas nesta pesquisa, nota-se que o conceito-chave de *research data* já era presente na literatura com 2016, já o termo *data literacy* aparece mais atrelado ao termo *Research Data Management* em 2017. O termo *Open Science* foi relacionado a partir de meados de 2016, e se mostra mais atrelado à ideia de *data sharing* em 2016. Isso significa que os artigos publicados nesse período começaram a aproximar o conceito de compartilhamento de dados ao conceito-chave de Ciência Aberta, evidenciando o interesse da comunidade científica nesta perspectiva de transparência dos dados gerando possibilidades de reúso.

Figura 16 - Relações de termos nas referências combinadas –WOS



Fonte: Autoria própria (2019).

A partir de cada estratégia de busca também foram selecionadas as teses e as dissertações mais relevantes, indicadas na Tabela 6, em um total de 21 trabalhos. Ao alterar a combinação de busca, os mesmos trabalhos podem ter sido recuperados em duplicidade pois atendem parcialmente os critérios da nova busca, considerando os anos de 2018 e 2019. A análise foi refeita em todos os resultados, sendo descartados os trabalhos repetidos, como detalhado na Tabela 6. As referências completas desses trabalhos estão listadas no Apêndice B.

Tabela 6 - Levantamento de teses e dissertações

Fontes/ Estratégias de Busca	OasisBR / BDTD	Catálogo de Teses da Capes	Selecionadas
Combinação A			
Dissertações	10	165	8 ³⁹
Teses	5	65	4 ⁴⁰
Combinação B			
Dissertações	32	0	0
Teses	13	0	0
Combinação C			
Dissertações	31	7	5 ⁴¹
Teses	12	4	4 ⁴²
Combinação D			
Dissertações	0	0	0
Teses	0	0	0
Combinação E			
Dissertações	8	85	0
Teses	2	38	0
TOTAL			
Dissertações	81	264	13
Teses	32	109	8
TOTAL GERAL	113	373	21

Fonte: Autoria própria (2019).

Nota-se que não há qualquer trabalho científico que considere a questão de Letramento Informacional em Dados Abertos de Pesquisa no contexto da Teoria Ator-Rede (combinação D) na Tabela 6, garantindo a originalidade deste tema combinado. No entanto, investigou-se também as outras combinações que incluíram os conceitos de Dados de Pesquisa (*Open Research Data*) e de Letramento Informacional em Dados (*Data information literacy*) pois são conceitos chaves deste estudo. Para tanto, buscou-se verificar nos trabalhos recuperados destas combinações, a partir de uma análise mais detalhada, a aderência ou não com o objeto do presente estudo.

Dentre todas as teses e dissertações recuperadas, apenas uma abordava o tema Dados de Pesquisa sob a perspectiva da Teoria Ator-Rede, porém, com o foco de propriedade intelectual na autoria científica a partir do compartilhamento de dados de pesquisa (MEDEIROS, 2015).

³⁹ APPEL, 2014; MACHADO, 2015; MONTEIRO, 2017; SOUZA, 2016; CARVALHO, (2018); FIGUEIREDO, (2018); JORGE, (2018); MAIA, (2018); SANTOS, (2018).

⁴⁰ MARQUES, 2014; MEDEIROS, 2015; PRADO, 2017.

⁴¹ OLIVEIRA, 2016; COELHO, 2017; AMORIM, 2013; NOVELLI, 2012; TYGEL, 2016.

⁴² SILVA, 2017; COSTA, 2017; ALVES, 2016; SOUSA, 2014.

Assim, ainda que haja mais trabalhos científicos abordando o tema Dados de Pesquisa e/ou Teoria Ator-Rede, os aspectos tratados não contemplam o letramento para reúso dos dados, o que reforça a originalidade dessa pesquisa. Apesar de terem sido recuperados 21 trabalhos, como detalhado no Quadro 10, apenas dois deles tiveram uma relação mais estreita com o objeto desta pesquisa para serem citados, são eles: o trabalho de Monteiro (2017), que trata de questões sobre direito autoral e propriedade intelectual em dados pesquisa e de Medeiros (2015), que aborda a propriedade intelectual na autoria científica sob a perspectiva da Teoria Ator-Rede.

Quadro 10 - Teses e dissertações selecionadas⁴³

Resultado de Levantamento de Teses e Dissertações	OasisBR / BDTD		Catálogo de Teses da Capes	
	Dissertação	Tese	Dissertação	Tese
A) Combinação A	10	5	133	57
Relevantes	CARVALHO, (2018); FIGUEIREDO, (2018); JORGE, (2018). MAIA, (2018); SANTOS, (2018);		4	3
			APPEL (2014); MACHADO (2015); MONTEIRO, (2017); SOUZA, (2016);	MARQUES, (2014); MEDEIROS, (2015); PRADO, (2017);
b) Combinação B	32	13	0	0
	Relevantes foram repetidos de outras estratégias			
c) Combinação C	31	12	1	0
Relevantes	OLIVEIRA, (2016); COELHO, (2017); AMORIM, (2013); NOVELLI, (2012); TYGEL, (2016).	SILVA, (2017); COSTA, (2017); ALVES, (2016); SOUSA, (2014).	Relevantes foram repetidos de outras estratégias	
d) Combinação D	0	0	0	0
e) Combinação E	8	2	52	30
Relevantes	0	0	Relevantes foram repetidos de outras estratégias	
f) Combinação F	0	0	1	0
Relevantes			Relevantes foram repetidos de outras estratégias	

Fonte: Autoria Própria (2019).

Ao observar-se as publicações científicas *stricto sensu* brasileiras distribuídas entre 2012 e 2019, nota-se que a maior parte (52%) é composta por dissertações, sendo a Ciência da Informação a área do conhecimento de maior concentração (62%) destes trabalhos. A maior

⁴³ A relação de referências completas das teses e dissertações pode ser vista no Apêndice A.

faixa de publicação está no ano de 2017 e 2018 (24% cada uma), o que sugere que o tema tem despertado interesse mais recentemente. A representatividade mais expressiva (29%) dos trabalhos está vinculada à Universidade Federal do Rio de Janeiro (UFRJ), que oferta curso de pós-graduação em Ciência da Informação. Esses dados estão apresentados na Tabela 7.

Tabela 7 - Publicações científicas *stricto sensu* brasileiras sobre gestão de dados de pesquisa – 2012-2019

Nível	qtde	%	Área de concentração	qtde	%	Ano	qtde	%	Univer-	qtde	%
									sidade		
Doutorado	10	48	Ciência da Computação / Informática	3	14	2012	1	5	UFF	1	5
Mestrado	11	52	Ciência da Informação	13	62	2013	1	5	UFGRS	3	14
						2014	4	19	UFPE	2	10
						2015	1	5	UFRJ	6	29
						2016	4	19	UFSCar	2	10
						2017	5	24	UNB	2	10
						2018	5	24	UNESP	2	10
								UNICA	1	5	
								MP			
								USP	2	10	
TOTAL	21	100		21	100		21	100		21	100

Fonte: Autoria Própria (2019).

A relação de referências completas das teses e dissertações pode ser vista no Quadro 30, no Apêndice A. Após a busca e recuperação dos artigos, teses e dissertações, procedeu-se a revisão sistemática da literatura para construir o *corpus* de análise da pesquisa.

5.5 PROCEDIMENTOS DE REVISÃO SISTEMÁTICA DA LITERATURA

Após a recuperação dos resultados no levantamento bibliográfico, os mesmos passaram por etapas de organização até serem efetivamente analisados de forma sistemática. A partir dos resultados das fontes pesquisadas, os registros foram exportados para pastas temporárias de cada uma das bases de dados pesquisadas. Essa seleção já excluiu os registros duplicados identificados pelo sistema de busca. Os registros marcados nas pastas temporárias foram exportados para um gerenciador de referências. O gerenciador utilizado foi o Zotero, ferramenta de uso gratuito *online* que também pode ser utilizada *off-line*, instalada no computador. A seleção dessa ferramenta se deu por oferecer diversas funcionalidades para a gestão das coleções, como por exemplo a organização e a localização dos registros por quaisquer metadados, a inclusão de textos completos como anexos em diferentes formatos, a geração de

listas de referências automáticas, de acordo com normas pré-selecionadas (ZOTERO, 2018). Após a organização das publicações, foi avaliada a duplicidade de registros provenientes de bases diferentes e a exclusão dos artigos repetidos. A partir da adequação dos registros é possível avançar para a seleção dos documentos mais pertinentes para a análise.

5.5.1 Seleção dos documentos - Etapa 3

Foi feita a leitura dos campos título, palavras-chave e resumo. A partir dessas informações os artigos foram selecionados para serem incluídos na análise sistemática, com os seguintes elementos identificados: (i) Teoria de Base, (ii) Objetivos, (iii) Palavras-Chave, (iv) Natureza da Pesquisa, (v) Abordagem metodológica, (vi) Procedimentos, (vii) Resultados e (viii) Problema. Todos esses itens foram identificados na planilha de apoio, como mostra o Quadro 11, exemplificando alguns artigos analisados. Após a descrição, o texto completo foi localizado e organizado em pastas virtuais. A leitura da introdução foi realizada e as produções com maior aderência foram selecionadas para leitura completa.

Quadro 11 - Amostra de artigos selecionados para a análise sistêmica

Autores	Ano	Título	Palavras-Chave	Objetivos	Natureza da Pesquisa	Abordagem metodológica	Procedimentos	Resultados	Problema
Koltay, Tibor	2015	Data literacy: in search of a name and identity	Research, Information Science, Library &, Academic Libraries, Data, Information Literacy, Document	The role of data literacy is discussed in the light of such activities as data quality, data management, data curation, and data citation. The differing terms and their relationship to the most important literacies are examined. The paper aims to discuss these issues	Teórica	The paper centres on information literacy, while the characteristics of other relevant literacies are also examined.	The content of data literacy education is explained in the context of data-related activities	Concluded that there is a need for data literacy and it is advantageous to have a unified terminology. Data literacy can be offered both to researchers, who need to become data literate science workers and have the goal to educate data management professionals. Several lists of competencies contain important skills and abilities, many of them indicating the close relationship between data literacy and information literacy. It is vital to take a critical stance on hopes and fears, related to the promises of widespread ability of (big) data	The importance of data literacy in fulfilling the mission of the contemporary
Tenopir, Carol, Elizabeth D. Dalton, Suzie Allard, Mike Frame, Ivanka Pjesivac, Ben Birch, Danielle Pollock, e Kristina Dorsett	2015	“Changes in Data Sharing and Data Reuse Practices and Perceptions among Scientists Worldwide”	Science, Q, Sciences (General), Medicine, R Collaboration, Research, Researchers, Studies, Information Sharing, Scientists, Open Access, Polls &, Surveys, Citation Indexes, Usability, National Science Foundation, Working Groups Medical Research – Information Management, Scientists – Information Management	Examine the current state of data sharing and reuse perceptions and practices among research scientists as they compare to the 2009/2010 baseline study, 2. to examine differences in practices and perceptions across age groups, geographic regions, and subject disciplines.	Empírico	Quantitativa	Surveys	Increased acceptance of and willingness to engage in data sharing, as well as an increase in actual data sharing behaviors. 2. However, there is also increased perceived risk associated with data sharing, and specific barriers to data sharing persist. 3. There are also differences across age groups, with younger respondents feeling more favorably toward data sharing and reuse, yet making less of their data available than older respondents. 4. Geographic differences exist as well, which can in part be understood in terms of collectivist and individualist cultural differences. 5. An examination of subject disciplines shows that the constraints and enablers of data sharing and reuse manifest differently across disciplines. 6. Implications of these findings include the continued need to build infrastructure that promotes data sharing while recognizing the needs of different research communities	The incorporation of data sharing into the research lifecycle is an important part of modern scholarly debate. to observe current states of data sharing and to see what, if any, changes have occurred in the past 3-4 years.

Fonte: Autoria própria (2017). Disponível em: <http://bit.ly/01-RSL>.

5.5.2 Extração e tratamento das informações para redação do trabalho – Etapas 4, 5 e 6

Depois de selecionadas as publicações, procedeu-se a leitura completa dos artigos, sendo realizado o fichamento de todo o material coletado, 80 documentos. As anotações foram sendo extraídas dos textos e o agrupamento das ideias que corroboravam, ou que se contrapunham, era constantemente avaliado, buscando coerência intelectual no trabalho. O intuito foi o de comparar e avaliar as informações obtidas a partir de fontes diferentes, que resulta em um recurso importante para explicar as ideias convergentes com o uso da citação agrupada entre parênteses, o que possibilita demonstrar as ideias que são corroboradas por diversos autores (ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS, 2002). Dessa forma, propostas/ideias/conceitos homogêneos, indicadas por distintos autores podem ser reforçadas pois apresentam um consenso dentro de uma comunidade.

Em seguida, procedeu-se a redação do trabalho, que apesar de estar estruturada de forma linear, ocorreu durante todo o procedimento de revisão sistemática. São as inferências a partir da leitura que trazem as condições de operacionalização da produção do referencial teórico e o trabalho pode ser produzido e reproduzido. Após a seleção e a análise dos artigos, procedeu-se com a seleção das teses e dissertações.

As teses e dissertações foram analisadas de acordo com o título a fim de se verificar a aderência do trabalho com os termos pesquisados, seguindo as mesmas etapas da análise sistemática. A análise foi refeita em todos os resultados sendo, posteriormente, descartados os trabalhos repetidos.

5.6 RESULTADO DA REVISÃO SISTEMÁTICA

A quantidade de documentos recuperados (total de 755) e as combinações selecionadas para análise encontram-se no Quadro 12. Ao julgar a eficácia dos resultados, notou-se que três autores detinham a maior quantidade de produção científica, portanto, em uma estratégia complementar foi realizada uma nova busca pelos nomes dos autores: Christine Borgman, Tilbor Koltay e Carol Tenopir. Esses resultados foram comparados com as estratégias combinadas e os artigos repetidos foram excluídos. Os demais foram analisados e quando pertinentes, adicionados ao *corpus* de análise. A quantidade dos registros recuperados, a distribuição por anos, a relação dos autores e seus artigos mais citados, além dos títulos desses artigos podem ser visualizadas no Apêndice C.

Quadro 12 - Combinação entre os descritores – quantidade recuperada

Busca Combinada	Idioma	Descritor 1	Descritor 2	Descritor 3	Capes	Ebscohost	Web of Science
Combinação A	Inglês	Open Science	Open Research Data		156	143	202
	Português	Ciência Aberta	Dados abertos de pesquisa		7	0	0
	Total				163	143	202
Combinação B	Inglês	Open Science	Open Research Data	Data information literacy	2	1	1
	Português	Ciência Aberta	Dados abertos de pesquisa	Letramento em dados	0	0	0
	Total				2	1	1
Combinação C	Inglês	Open Research Data	Data information literacy		19	116	38
	Português	Dados abertos de pesquisa	Letramento em dados		0	0	0
	Total: 173				19	116	38
Combinação D	Inglês	Open Research Data	Data information literacy	Actor-Network theory	6	0	0
	Português	Dados abertos de pesquisa	Letramento em dados	Teoria Ator-Rede	0	0	0
	Total				6	0	0
Combinação E	Inglês	Virtual Research Environment	Open Research Data		4	24	36
	Português	Ambiente Virtual de Pesquisa	Dados abertos de pesquisa		0	0	0
	Total: 64				4	24	36
Total analisado: 237					194	284	277
						755	

Fonte: Autoria própria (2019).

A seleção dos resultados que compuseram o *corpus* de análise foram as estratégias das combinações C e E, apresentadas na Tabela 8. A estratégia A trouxe um volume maior de referências, porém, ao realizar a leitura preliminar, percebeu-se que os resultados eram bem abrangentes, tratando de distintas abordagens da Ciência Aberta e dos Dados de Pesquisa, não necessariamente aderentes ao foco deste estudo.

Destaca-se que a combinação B já estava implícita na combinação C. A combinação C

demonstrou ser a mais aderente, pois aborda o objeto de pesquisa (dados abertos) sob a perspectiva do letramento, foco de estudo. A combinação D apesar de recuperar resultados, trouxe referências duplicadas da combinação C.

A Combinação E foi selecionada pela necessidade de ambientação com o termo “ambiente virtual de pesquisa”. Ao final foram selecionados os 173 artigos provenientes da combinação C e os 64 artigos da Combinação E, totalizando 237 artigos que constituíram o *corpus* preliminar de análise, como detalhado na Tabela 8, por base pesquisada, apresentando o percentual de recuperação de cada base e a seleção em cada uma delas.

Tabela 8 - Seleção das referências para compor *corpus* de análise – busca por descritores – 2019

Busca Combinada	Capes	Ebsco Host	Web of Science	Total	% por combinação	N Selecionadas	% selecionadas
Combinação A	163	143	202	508	67,28%	0	0,00%
Combinação B	2	1	1	4	0,53%	0	0,00%
Combinação C	19	116	38	173	22,91%	173	73,00%
Combinação D	6	0	0	6	0,79%	0	0,00%
Combinação E	4	24	36	64	8,48%	64	27,00%
Total referências	194	284	277	755	100,00%	237	38,20%
% de referências por bases	25,26 %	38,86%	35,88%	100,00 %			

Fonte: Autoria própria (2019).

A análise foi selecionada com base nos resultados das combinações C e E (287 artigos), além da pesquisa complementar pelos autores mais recorrentes encontrados nesses resultados para Borgman, Koltay e Tenopir. A recuperação da busca adicional pelos autores está detalhada na Tabela 9.

Tabela 9 - Seleção das referências para compor *corpus* de análise – busca por autores

AUTORES	Capes	Ebsco Host	Web of Science	Total	% por AUTORES	N Seleccionadas	% seleccionadas
BORGMAN	19	17	18	54	18,82%	54	18,82%
KOLTAY	27	37	12	76	26,48%	76	26,48%
TENOPIR	42	85	30	157	54,70%	157	54,70%
Total referências	88	139	60	287	100,00%	287	100,00%
% de referências por bases	30,66%	48,43%	20,91%	100,00%			

Fonte: Autoria própria (2019).

A seguir são apresentados as etapas dos procedimentos para a análise das 237 publicações (resultados das buscas combinadas) e das 287 publicações (resultados das buscas por autores), totalizando 524 referências selecionadas a fim de identificar as que seriam aderentes ao objeto de pesquisa. A partir da leitura do título, das palavras chaves e do resumo, as etapas a seguir foram executadas a fim de se estabelecer o *corpus* de análise final.

5.6.1 Etapas da Revisão Sistemática de Literatura (RSL)

Partiu-se da análise prévia de 524 referências encontradas que passaram por cinco etapas como: (i) pesquisa e recuperação, (ii) eliminação de duplicadas e exportação, (iii) seleção pela verificação do título e do resumo até serem (iv) incluídos na lista de seleção para leitura completa sendo localizados os (v) textos completos disponíveis. Ao final foram selecionados 80 resultados aderentes à pesquisa, a serem incorporados na análise sistemática. O resultado numérico resumido dos documentos recuperados em todas as etapas descritas anteriormente é apresentado na Tabela 10. Os dados numéricos completos de cada uma das etapas, desde a recuperação nas bases de dados até a etapa de leitura completa dos 80 artigos selecionados, estão disponíveis na Tabela 20 no Apêndice C.

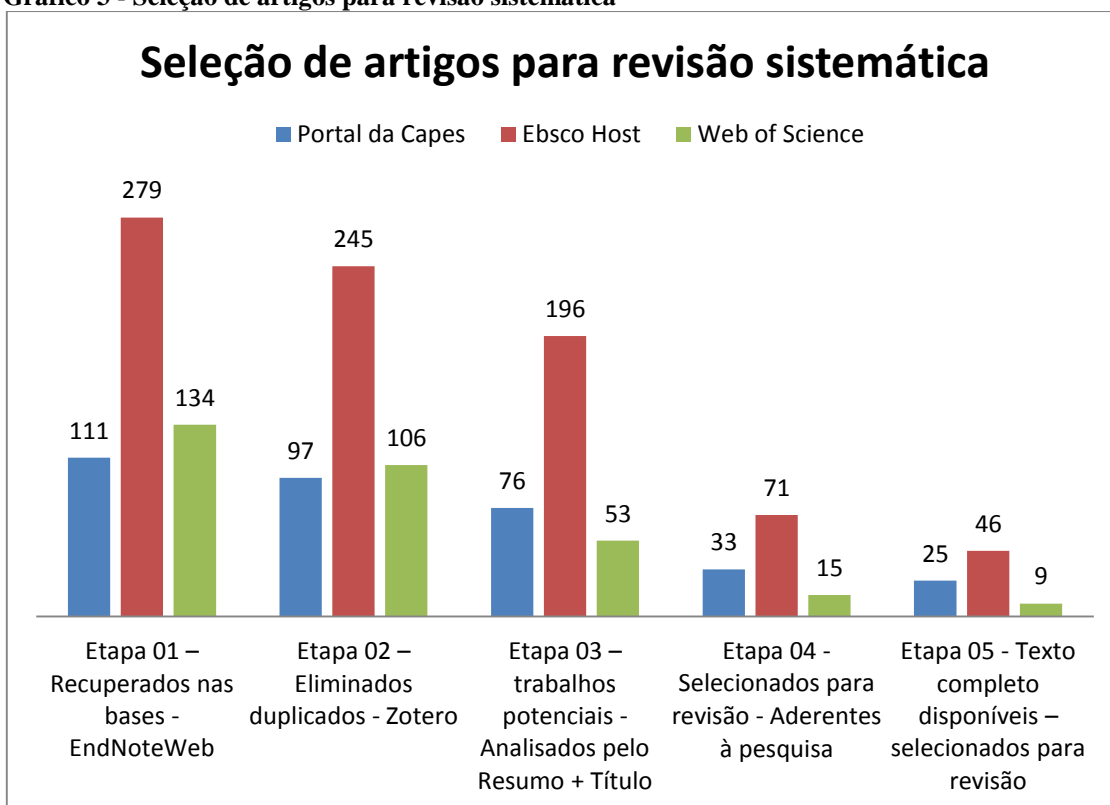
Tabela 10 - Resumo da seleção final de artigos para revisão sistemática – busca por descritores + autores

Estratégia de Pesquisa	Portal da Capes		Ebsco Host		Web of Science		Total
	Número	%	Número	%	Número	%	
TOTAL							
Etapa 01 – Recuperados nas bases - EndNoteWeb	111	21%	279	53%	134	26%	524
Etapa 02 – Eliminados duplicados - Zotero	97	22%	245	55%	106	24%	448
Etapa 03 – trabalhos potenciais - Analisados pelo Resumo + Título	76	23%	196	60%	53	16%	325
Etapa 04 - Seleccionados para revisão - Aderentes à pesquisa	33	28%	71	60%	15	13%	119
Etapa 05 - Texto completo disponíveis – selecionados para revisão	25	31%	46	58%	9	11%	80
TOTAL referências Seleccionadas para análise do texto completo	Portal da Capes		Ebsco Host		Web of Science		Total
	Número	%	Número	%	Número	%	
	25	31%	46	58%	9	11%	80

Fonte: Autoria própria (2019).

As quantidades de publicações avaliadas em cada etapa, com os dados da Tabela 10 são apresentadas no Gráfico 3. A evolução dos filtros em cada etapa de análise é demonstrada para as referências localizadas em cada uma das bases de dados pesquisadas: Portal da Capes, Ebsco Host e Web of Science. O Gráfico 3 demonstra que a maior parte das referências foram localizadas no portal Ebsco Host, iniciando com 279 referências e sendo selecionadas 46 para leitura do texto completo. No portal da Capes a primeira etapa contou com 111 referências e finalizou com 25. Na Web of Science foram recuperadas 134 referências e apenas 9 foram selecionadas para leitura do texto completo, principalmente por serem duplicações das referências do portal Ebsco Host.

Gráfico 3 - Seleção de artigos para revisão sistemática



Fonte: Autoria própria (2019).

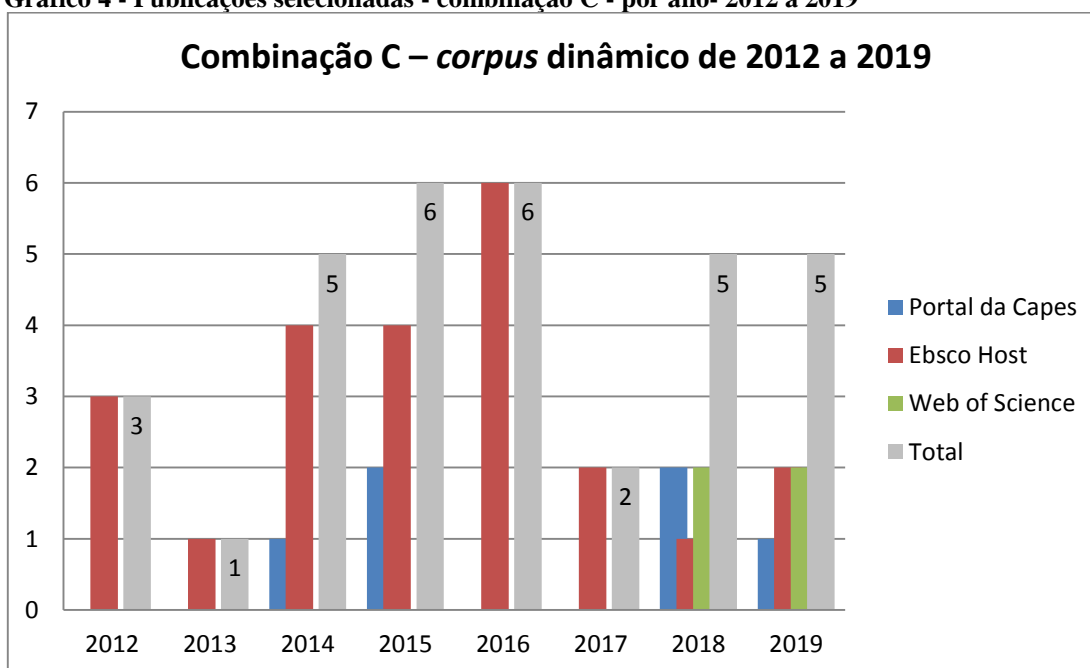
As referências selecionadas em cada base de dados foram distribuídas por anos, conforme a Tabela 11 e a Tabela 12. Apesar de a pesquisa ter sido delimitada nos últimos 5 anos, alguns motores de busca recuperaram outros anos. Os artigos fora do período de delimitação não foram incluídos na análise sistemática. A distribuição da quantidade de registros selecionados na estratégia selecionada Combinação C, pode ser vista na Tabela 11, que demonstra a distribuição por ano, iniciando em 2012 até 2019, separada por bases de dados pesquisadas. A maior quantidade recuperada foi da combinação C, proveniente do portal Ebsco Host referente ao ano de 2016, base que apresentou 70% dos resultados da combinação C. No total foram analisados na revisão sistemática 33 artigos referentes à combinação C.

Tabela 11 - Combinação C – corpus dinâmico de 2012 a 2019

Combinação C	Portal da Capes	Ebsco Host	Web of Science
2012	0	3	0
2013	0	1	0
2014	1	4	0
2015	2	4	0
2016	0	6	0
2017	0	2	0
2018	2	1	2
2019	1	2	2
TOTAL	6	23	4
%	18%	70%	12%
TOTAL C		33	

Fonte: Autoria própria (2019).

A evolução temporal de produção dos registros selecionados na estratégia de busca Combinação C, pode ser vista no Gráfico 4. A maior concentração está no portal EbscoHost no ano de 2015 e 2016 (6 referências cada ano), conforme já mencionado, seguida pelos anos de 2014, 2018 e 2019 (5 referências cada ano), 2012 (3 referências), 2017, (2 referências) e 2013 (1 referência).

Gráfico 4 - Publicações selecionadas - combinação C - por ano- 2012 a 2019

Fonte: Autoria própria (2019).

A distribuição da quantidade de registros selecionados na estratégia de busca Combinação E pode ser vista na Tabela 12. As referências recuperadas no portal EbscoHost e do portal da Capes são bem semelhantes, com apenas um registro de diferença. A distribuição cronológica também não apresentou variações expressivas. No total foram 17 referências analisadas da combinação E.

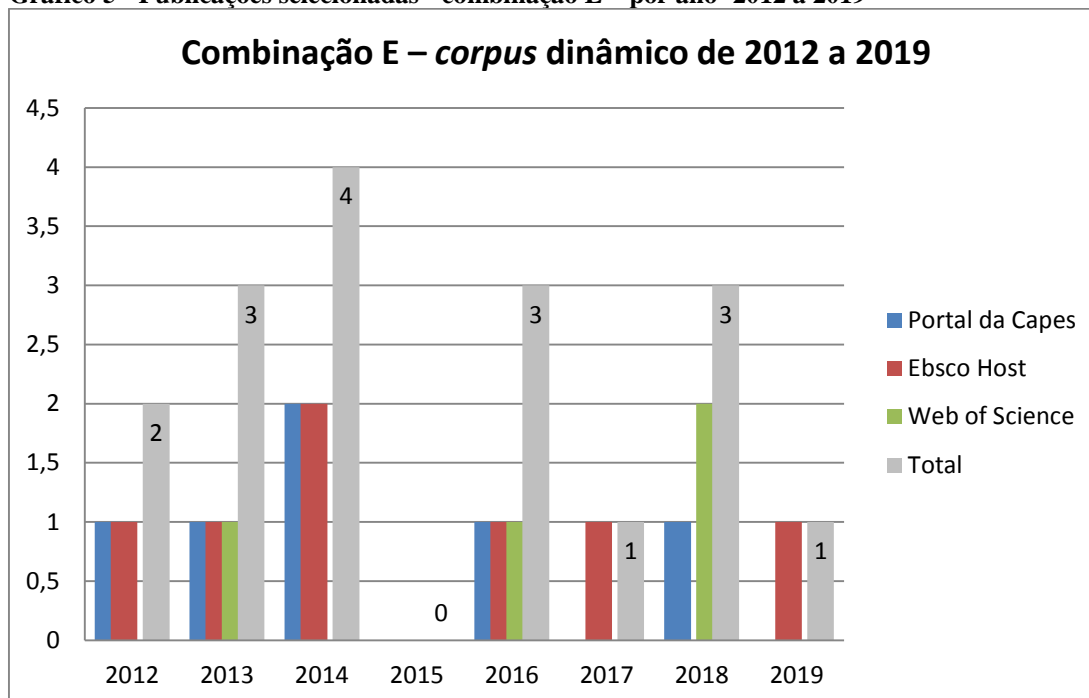
Tabela 12 - Combinação E – *corpus* dinâmico de 2012 a 2019

Combinação E	Portal da Capes	Ebsco Host	Web of Science
2012	1	1	0
2013	1	1	1
2014	2	2	0
2015	0	0	0
2016	1	1	1
2017	0	1	0
2018	1	0	2
2019	0	1	0
TOTAL	6	7	4
%	35%	41%	24%
TOTAL E		17	

Fonte: Autoria própria (2019).

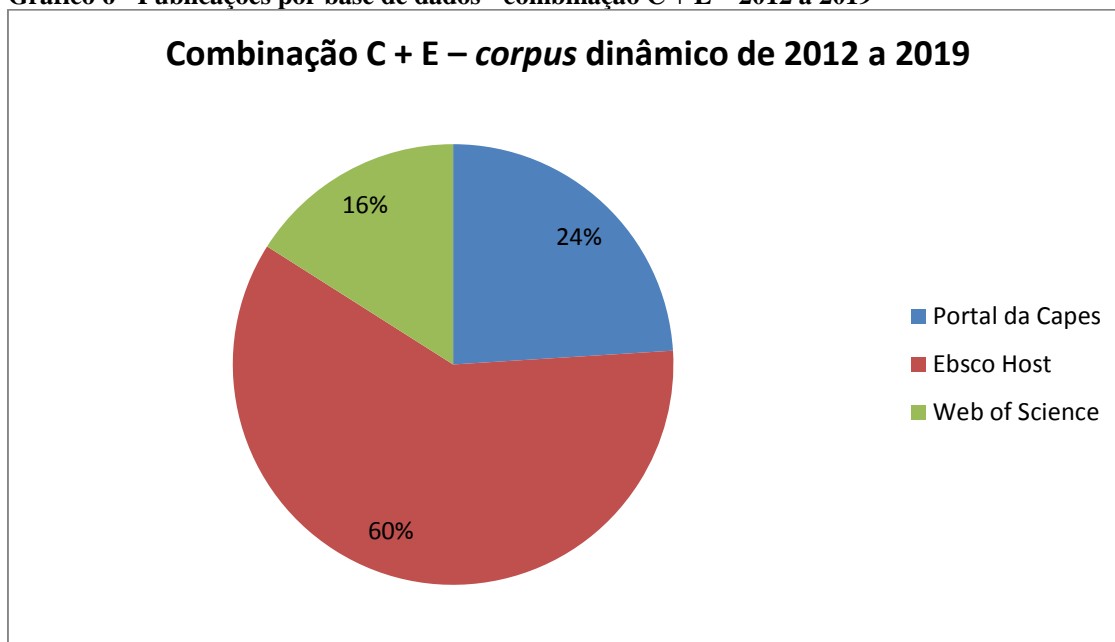
A evolução cronológica de produção dos registros recuperados na estratégia de busca Combinação E, pode ser vista no Gráfico 5. Nota-se que nenhuma referência foi recuperada no ano de 2015. A quantidade de referências recuperadas no portal da Capes e no portal EbscoHost são equivalentes para o ano de 2014 e 2012. Todas as bases pesquisadas apresentam pelo menos uma referência cada uma, nos anos de 2013 e 2019.

Gráfico 5 - Publicações selecionadas - combinação E – por ano- 2012 a 2019



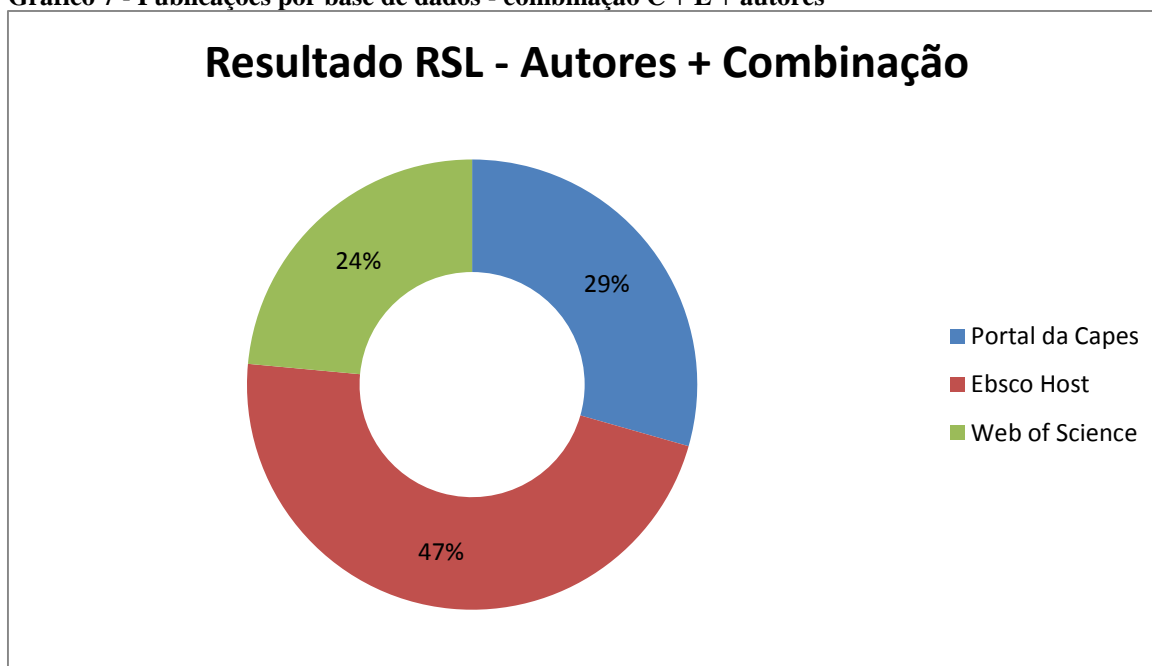
Fonte: Autoria própria (2019).

A proporção dos registros selecionados na estratégia de busca C + E por bases de dados pode ser vista no Gráfico 6. Nota-se o maior resultado recuperado no portal EbscoHost, principalmente devido às bases de dados específicas da área da Ciência da Informação e Biblioteconomia disponíveis nesse portal, como a Library, Information Science & Technology Abstracts with Full Text (LISTA), e a Information Science & Technology Abstracts (ISTA). Isso reforça a assertividade da seleção da fonte de informação para a pesquisa.

Gráfico 6 - Publicações por base de dados - combinação C + E – 2012 a 2019

Fonte: Autoria própria (2019).

Ao analisar a proporção de todas as publicações selecionadas para compor o *corpus* da análise da Revisão Sistemática de Literatura (RSL), por bases de dados, nota-se a predominância dos resultados recuperados no portal Ebsco Host, com 49%, seguido do portal da Capes, com 26% e por último a Web of Science, com 25%, como pode ser visto no Gráfico 7.

Gráfico 7 - Publicações por base de dados - combinação C + E + autores

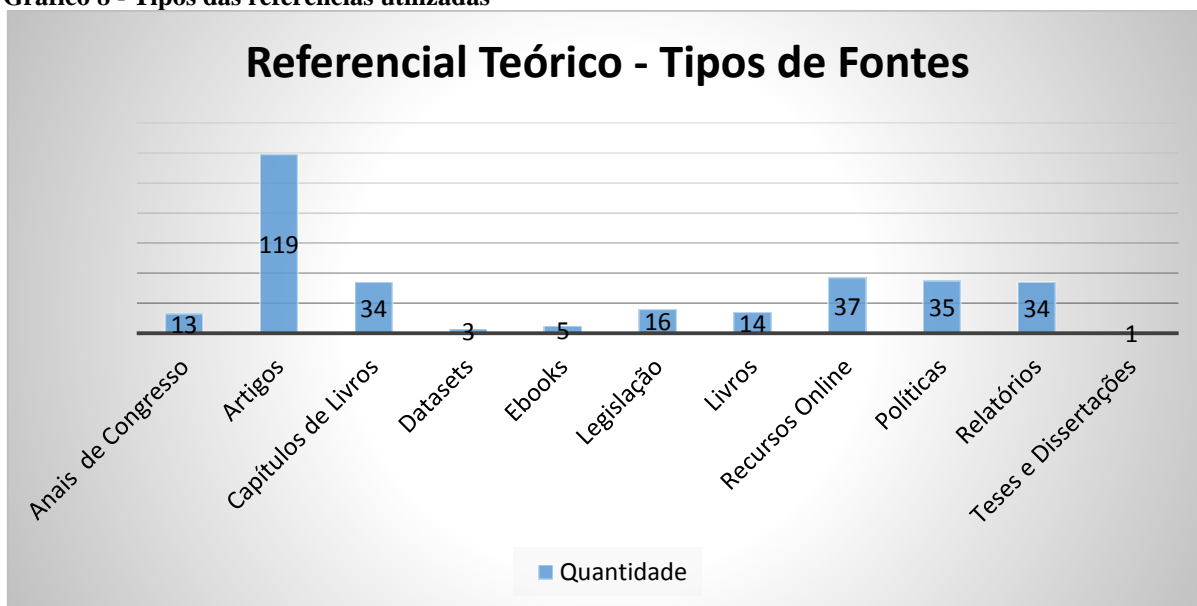
Fonte: Autoria própria (2019).

Após a seleção e a análise estatística da recuperação de resultados das referências, seguiu-se para a extração das informações e elaboração do texto. Após os levantamentos bibliográficos de artigos, de teses e de dissertações, uma análise do *corpus* teórico dinâmico utilizado é apresentada na próxima seção, com a consolidação da Revisão Sistemática.

5.6.2 Consolidação da Revisão Sistemática: *Corpus* teórico utilizado

Considerando contribuir com uma visão sistêmica do *corpus* referencial utilizado nessa pesquisa, apresenta-se algumas análises sobre as referências utilizadas. Todas as 314⁴⁴ referências utilizadas (citadas) – Revisão de Literatura, somados os *corpora* estático e dinâmico foram classificadas quanto ao assunto principal, ao formato e a data de publicação. A composição por tipo de formato para os documentos e as quantidades de cada formato é apresentada no Gráfico 8.

Gráfico 8 - Tipos das referências utilizadas

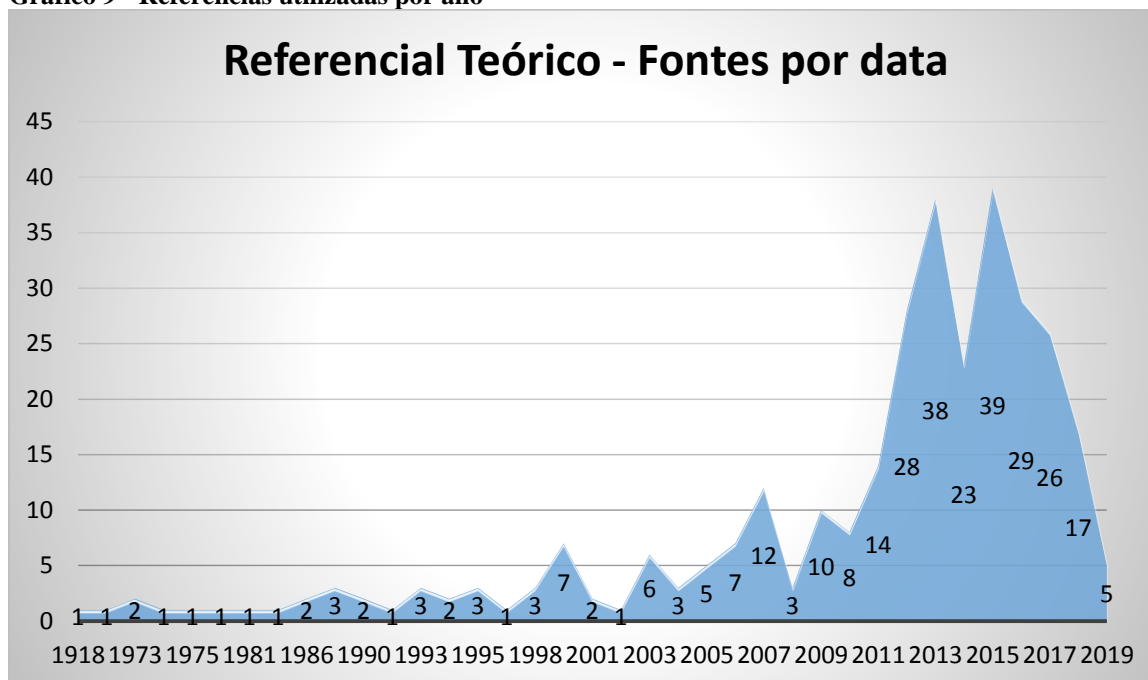


Fonte: Autoria própria (2019).

A distribuição cronológica das referências utilizadas é mostrada Gráfico 9, sendo as referências distribuídas entre 1918 e 2019.

⁴⁴ Número até a finalização das estatísticas, podendo ser adicionadas novas referências.

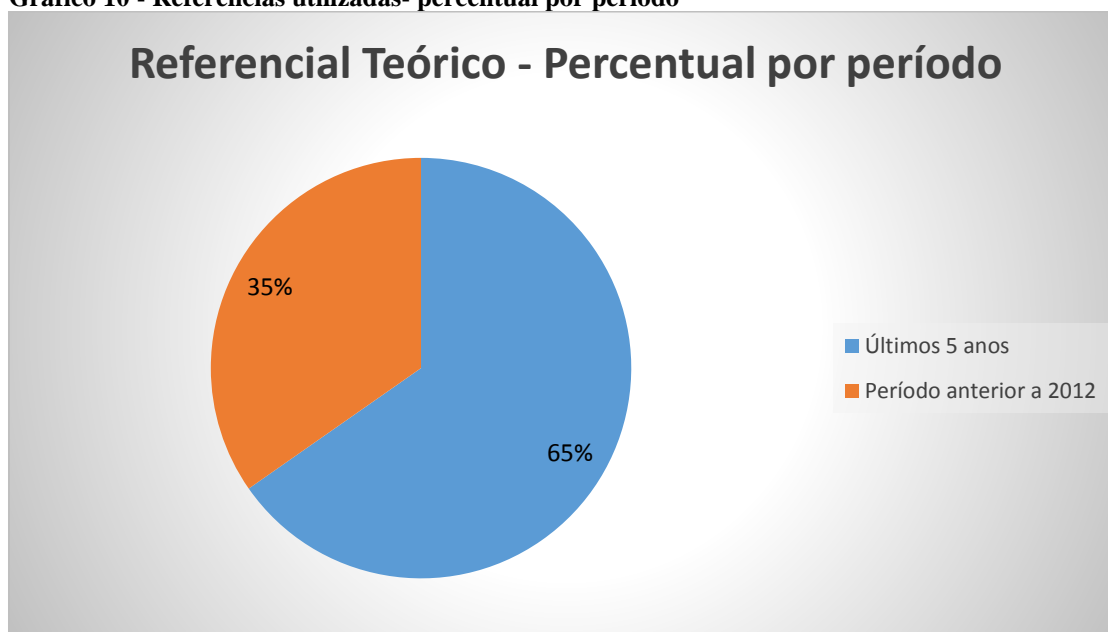
Gráfico 9 - Referências utilizadas por ano



Fonte: Autoria própria (2019).

Das 314 referências citadas (*corpus* estático e *corpus* dinâmico), 205 (65%) referem-se aos últimos 5 anos, sendo que 109 referem-se a períodos anteriores (35%), como apresenta-se no Gráfico 10.

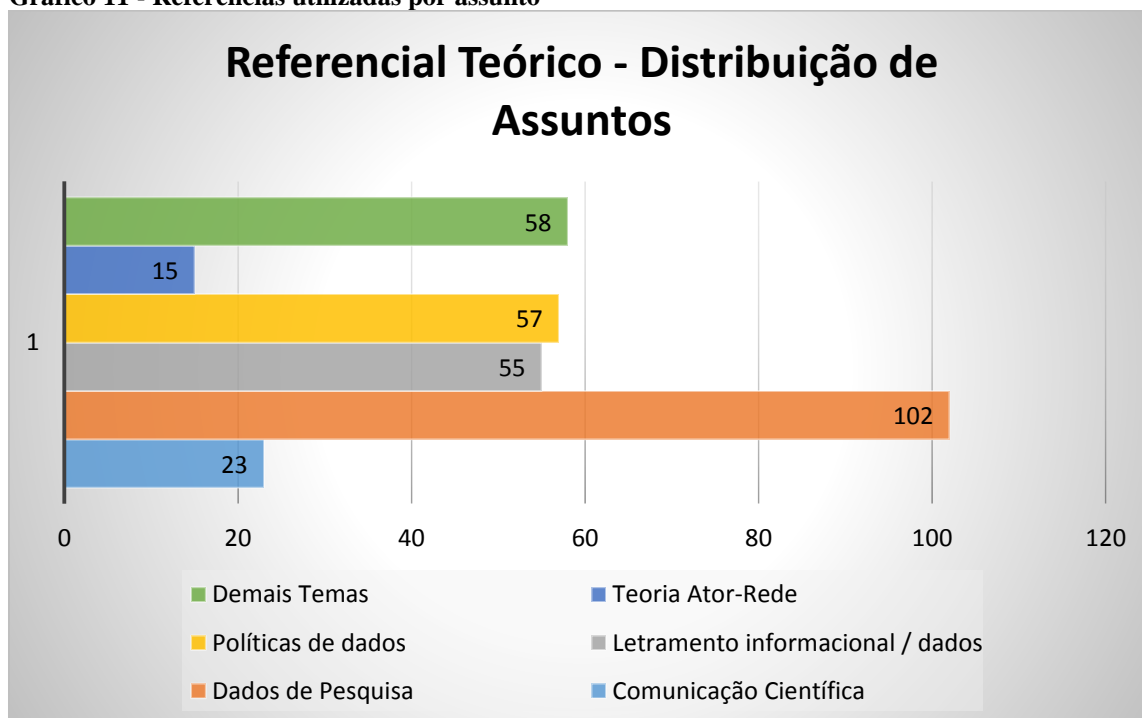
Gráfico 10 - Referências utilizadas- percentual por período



Fonte: Autoria própria (2019).

Em relação aos temas cobertos pela literatura utilizada, as quantidades por temas, dentre os mais abordados, sendo 102 relacionados aos dados de pesquisa, 57 com as questões de políticas de dados, 55 voltados para o embasamento de letramento informacional, 23 referências da Comunicação científica, 15 em relação à Teoria Ator-rede e os demais 58 itens relacionados a outros temas para embasar o trabalho são apresentadas no Gráfico 11.

Gráfico 11 - Referências utilizadas por assunto



Fonte: Autoria própria (2019).

Após a demonstração de como foram selecionados os documentos para a elaboração do *corpus* teórico da pesquisa, apresenta-se a metodologia de análise qualitativa do referencial identificado na literatura. A partir da visão da cobertura que se pretendeu dar do tema e das suas intersecções mediante o *corpus* teórico utilizado, propõe-se o protocolo para a execução da pesquisa. A partir da revisão sistemática da literatura foi possível identificar e selecionar os modelos conceituais (*frameworks*) sobre letramento em dados, cujo resultado da análise é descrito a seguir.

5.7 METODOLOGIA DE ANÁLISE DE DADOS QUALITATIVOS - PROTOCOLO DE EXECUÇÃO

O Protocolo de execução apresenta como foram coletados e tratados os dados para atingir os objetivos específicos, propondo as técnicas e os procedimentos empregados na execução de cada um e os resultados obtidos que, a partir de uma análise em triangulação, viabilizaram a constituição de um modelo de referência com os requisitos e as competências para o reúso de dados de pesquisa em Ciências Sociais.

O que se pretende com a triangulação é a garantia da sinergia entre as características comuns dos conteúdos e dos dados analisados, organizando as competências para o letramento que permitam o reúso de dados. As características serão validadas seguindo as recomendações de boas práticas para dados disponíveis *online*, propostos pelo World Wide Web Consortium (W3C) (LÓSCIO; BURLE; CALEGARI, 2017). Esse é um importante instrumento que aborda 35 boas práticas de gestão de dados na internet, disponível no Apêndice R. O reúso de dados é o único benefício listado que se aplica às 35 recomendações, portanto, são práticas adequadas para atestar se os requisitos levantados estão aderentes à recomendação da W3C.

Os objetivos, o locus de coleta, o tratamento metodológico, os procedimentos de análise, as ferramentas e os resultados esperados, indicando a seção na qual serão descritos, estão apresentados no Quadro 13.

Quadro 13 - Protocolo geral de execução de pesquisa

(continua)

Conteúdo/Objetivos	Lócus de Coleta	Tratamento metodológico	Procedimentos de análise	Ferramentas de análise	Resultados
Tema: dados abertos de pesquisa e o letramento para reúso	Revisão de Literatura	Qualitativo	Levantamento Bibliográfico e Revisão Sistemática de Literatura (RSL)	<i>Software</i> Zotero para organização das referências	Capítulo 1
Questão de Pesquisa: Quais são os requisitos necessários em ambientes virtuais de aprendizagem para que os pesquisadores desenvolvam competências mínimas de letramento informacional para reúso de dados de pesquisa nas Ciências Sociais, na abordagem da Teoria Ator-Rede?	Revisão de Literatura	Qualitativo	Resultados dos objetivos Revisão de Literatura	<i>Software</i> Zotero para organização das referências	Capítulo 6
Objetivo Específico 1 – OE1 Caracterizar ambientes virtuais de pesquisa formados pela infraestrutura de coleta, de armazenamento, de serviços, de disponibilidade e de reúso de dados de pesquisa com base na literatura	Revisão de Literatura	Qualitativo	RSL Análise de conteúdo	Planilha eletrônica para desdobramento do conteúdo das referências <i>Software</i> NVivo	Capítulos 2 e 3
Objetivo Específico 2 - OE2 Levantar nos quadros conceituais (<i>frameworks</i>) de letramento informacional em dados, quais são os requisitos propostos para o reúso de dados;	<i>Frameworks</i> Revisão de Literatura	Qualitativo	RSL Análise de Conteúdo Categorias de análise	<i>Software</i> NVivo Gráficos Frequência de palavras	Capítulo 5 Seção 5.9.1 Quadro 19
Objetivo Específico 3 – OE3 Identificar as perspectivas e as dificuldades dos pesquisadores no reúso dados de pesquisa, por meio de dados secundários (<i>datasets</i>);	Dados secundários Revisão de Literatura	Quantitativo / Qualitativo	Reúso de dados Seleção de campos pertinentes Contagem de repostas Análise de Conteúdo Categorias de contexto	<i>Software</i> Excel Seleção de campos Seleção de Filtros Contagem de frequência	Capítulo 6 Seção 6.1
Objetivo Específico 4 – OE4 Identificar as competências essenciais para o reúso de dados no letramento proposto pelos recursos de aprendizado online, disponíveis nos ambientes virtuais de aprendizagem;	Recursos de aprendizagem <i>online</i> Revisão de Literatura	Qualitativo	Análise de conteúdo Unidades de contexto	<i>Software</i> NVivo Gráficos Codificação	Capítulo 6 Seção 6.2 Quadro 24
Objetivo Específico 5 – OE5 Comparar as características dos quadros conceituais com as perspectivas dos pesquisadores e com as competências nos recursos de aprendizagem online, propondo um conjunto de requisitos com vistas ao reúso de dados nas Ciências Sociais;	Comparativo das análises dos OE2, OE3 e OE4 Revisão de Literatura	Qualitativo	Matrizes das Análise de conteúdo Triangulação de dados	<i>Software</i> NVivo Gráficos Codificação	Capítulo 6 Seção 6.3 Quadro 25

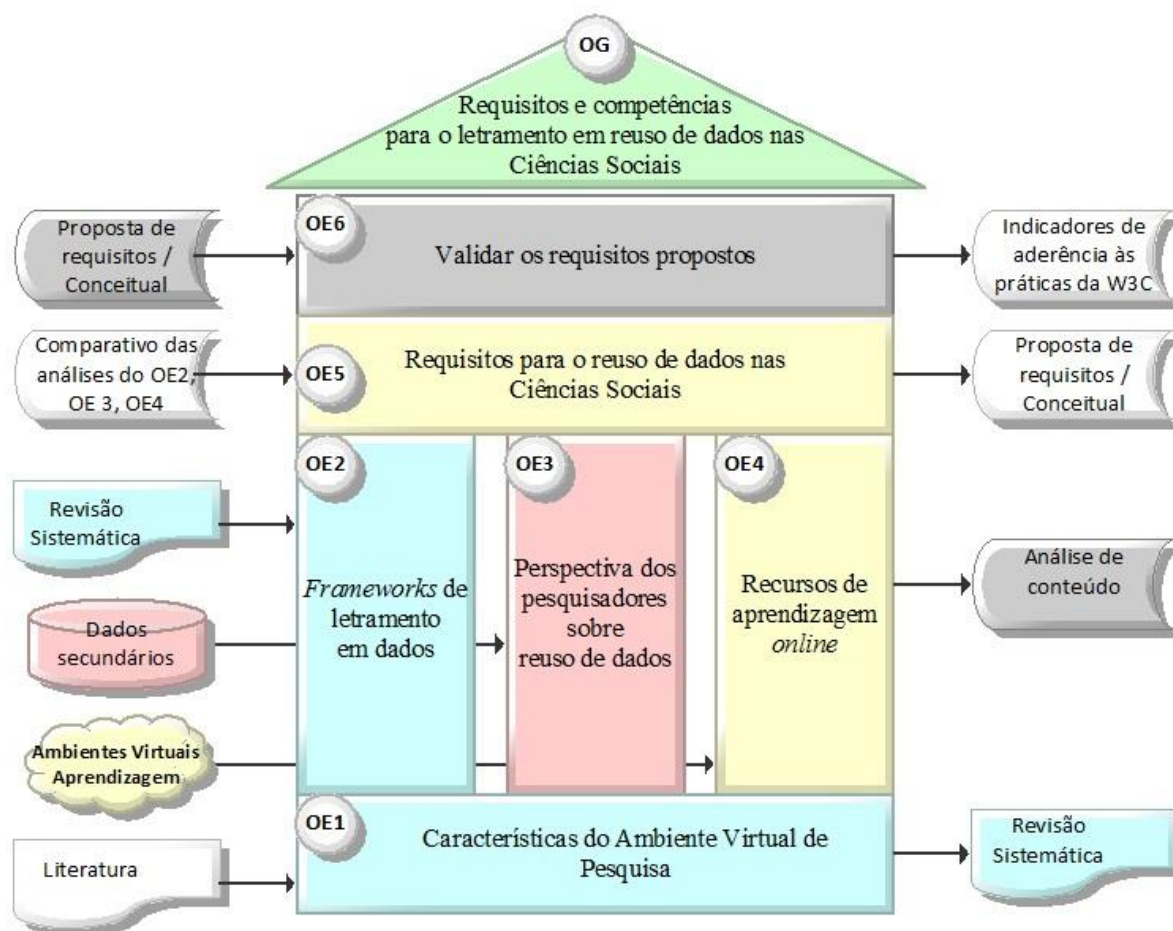
(conclusão)

Conteúdo/Objetivos	Lócus de Coleta	Tratamento metodológico	Procedimentos de análise	Ferramentas de análise	Resultados
Objetivo Específico 6 – OE6 Validar os requisitos propostos por meio da verificação de aderência às melhores práticas da área recomendadas pelo World Wide Web Consortium (W3C).	Validação: aderência do OE5 com as melhores práticas de Gestão de Dados da W3C	Quantitativo / Qualitativo	Percentual de aderência às melhores práticas da <i>Web</i>	Planilha eletrônica com Indicadores e Gráficos	Capítulo 6 Seção 6.5 Quadro 28 Gráfico 20
Objetivo Geral Propor, a partir de elementos da Teoria Ator-Rede, requisitos para um ambiente virtual em letramento informacional de dados, com as competências mínimas para o reuso de dados em Ciências Sociais.	Requisitos do OE6 Revisão de Literatura	Qualitativo	Análise de conteúdo Proposta Conceitual	<i>Software</i> NVivo	Capítulo 6 Seção 6.4

Fonte: Autoria própria (2018).

O objetivo específico 1 (OE1) - Características do Ambiente Virtual de Pesquisa, utilizou como literatura fontes para que, a partir da revisão sistemática da literatura, os outros próximos três objetivos específicos pudessem ser alcançados. O objetivo específico 2 (OE2) utilizou a revisão sistemática da literatura para identificar os *Frameworks* de letramento em dados que, depois de identificados, foram tratados sob a perspectiva de análise de conteúdo. O objetivo específico 3 (OE3) utilizou dados secundários (*datasets*) para identificar as perspectivas e as dificuldades dos pesquisadores no reuso de dados. O objetivo específico 4 (OE4) buscou a partir dos Ambientes Virtuais de Aprendizagem (AVAs) os Recursos de aprendizagem *online* disponíveis para que por meio da análise de conteúdo fossem identificadas as competências e as habilidades abordadas. A análise comparativa dos resultados dos OE2, OE3 e OE4 constituiu o objetivo específico 5 (OE5) que é o de comparar as características dos quadros conceituais com as perspectivas dos pesquisadores e com as competências nos recursos de aprendizagem *online*, propondo um conjunto de requisitos com vistas ao reuso de dados nas Ciências Sociais. A partir das características identificadas, finalizou-se com o objetivo específico 6 (OE6), em que os requisitos propostos decorrente dessas análises foram validados, testando o seu grau de aderência às recomendações das melhores práticas em Gestão de dados estabelecidas pela W3C, via análise de conteúdo a partir dos estudos de Lóscio, Burle e Calegari (2017). As descrições das melhores práticas foram avaliadas de forma a identificar qual ou quais requisitos eram compatíveis, por meio de análise semântica e léxica das palavras. As inter-relações dos objetivos específicos entre si e com o objetivo geral estão exemplificadas na Figura 17.

Figura 17 - Inter-relação dos objetivos específicos e geral



Fonte: Autoria própria (2018).

Ao procurar responder à questão de pesquisa inicial, juntamente com os objetivos desdobrados, foram identificados os conjuntos de dados, os quadros conceituais (*frameworks*) e os recursos de aprendizagem *online*, analisados posteriormente. Os procedimentos de coleta e de tratamento dos dados serão apresentados na próxima seção.

5.8 PROCEDIMENTOS DE COLETA E DE ANÁLISE DE DADOS SECUNDÁRIOS

Esta pesquisa teve como principais fontes de dados: (i) a revisão sistemática da Literatura (*frameworks*); (ii) os conjuntos de dados para reuso (*datasets*) e (iii) os recursos de aprendizagem disponíveis *online*.

A coleta de dados secundários ocorreu em bases de repositórios institucionais que disponibilizam dados de pesquisa em acesso aberto. A forma de localização dos repositórios se

deu a partir de uma pesquisa no indexador re3data.org, que lista os repositórios de dados registrados em seu sistema. Um dos importantes repositórios na área das Ciências Sociais é o repositório Inter-university Consortium for Political and Social Research (ICPSR), estabelecido em 1962, reúne mais de 750 instituições acadêmicas e organizações de pesquisa, mantém mais de 250.000 arquivos de dados e é colaborador de diversas agências de fomento (INTER-UNIVERSITY CONSORTIUM FOR POLITICAL AND SOCIAL RESEARCH, 2019). Nesta fonte localizaram-se dados de pesquisa de comportamentos dos pesquisadores em relação aos reuso de dados.

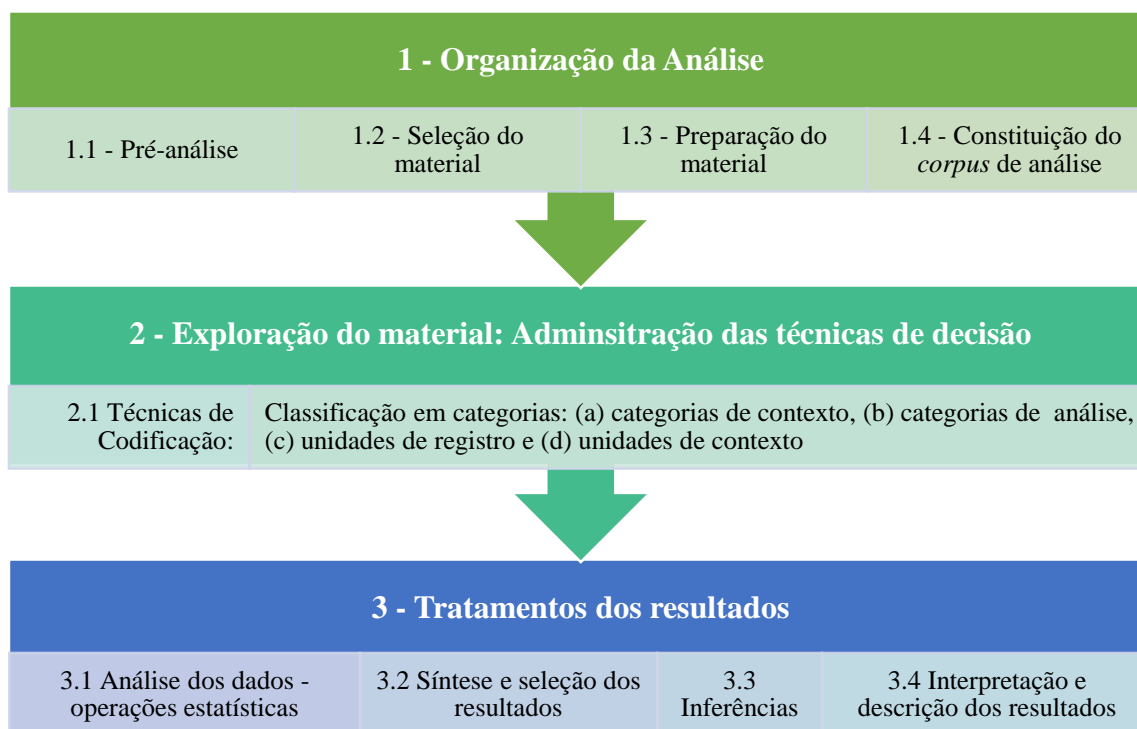
Usou-se, igualmente, a triangulação de dados na análise, buscando comparar os resultados das análises da literatura (*frameworks*) com os dados secundários de pesquisa e os recursos de aprendizagem em gestão de dados disponíveis no Quadro 4. Dessa forma, as práticas educacionais dos ambientes virtuais de aprendizagem voltados aos dados de pesquisa também foram investigadas do ponto de vista dos requisitos e competências essenciais para o reuso de dados. A técnica empregada foi a análise de conteúdo, cujo detalhamento segue na próxima seção.

5.8.1 As técnicas de análise – análise de conteúdo e de dados

Como técnica de tratamento dos dados para a análise dos quadros conceituais (*frameworks*) e dos dados de pesquisa secundários coletados, este trabalho utilizou a técnica de análise de conteúdo. A análise de conteúdo é considerada, para Bardin (1995) uma técnica empírica, que segue uma série de prerrogativas que devem ser consistentes ao longo de toda análise: desde a (i) busca exaustiva de conteúdo representativo, homogêneo e pertinente passando pelas (ii) análises quali-quantitativas, utilizando as (iii) interpretações das evidências levantadas para a (iv) descrição das inferências alcançadas

Foi utilizado o *software* NVivo para apoiar a fase de análise dos resultados, para as etapas de (a) codificação e de (b) classificação das categorias. A codificação é a etapa na qual os dados são categorizados, ou seja, recebem atributos que os qualificam por uma determinada característica. Já a classificação em categorias envolve a organização e a ordenação de uma série de dados em diferentes classes, a partir de uma ou mais variáveis. As etapas do método de análise de conteúdo aplicadas nesta pesquisa são ilustradas na Figura 18.

Figura 18 - Etapas da análise de conteúdo



Fonte: Adaptado de Bardin (1995, p. 102).

Bardin (1995) identifica algumas vantagens no tratamento informático da análise de conteúdo, como por exemplo: aumento do rigor pelo controle das operações, maior rapidez e melhoria da manipulação de dados complexos e aumento da facilidade de reprodução e troca de documentos.

Para as análises dos dados, utilizou-se de diversos recursos de codificação e visualização da informação na pré-análise. Utilizou-se também de técnicas de codificação automatizada com apoio do *software* NVivo. A codificação foi realizada a partir dos arquivos dos *frameworks* em formato *Portable Document Format* (PDF) a fim de, a partir da análise do seu conteúdo, identificar as categorias de contexto e as categorias de análise. A organização das categorias de análise e as unidades de registros iniciais são apresentadas nas descrições das etapas de análise de conteúdo, sequencialmente.

A categorização, segundo Bardin (1995), visa a organização de uma série de dados em diferentes classes, agrupados por características comuns, a partir de uma análise da terminologia empregada para representar os conceitos. A categorização foi criada e hierarquizada de acordo com os princípios norteadores de Bardin (1995):

a) categorias de Contexto: categorias mais abrangentes, orientadas pelo levantamento bibliográfico;

- b) categorias de Análise: segmentação das categorias de contexto;
- c) unidades de Registro: subdivisões da categoria de análise;
- d) unidades de Contexto: detalhamento das unidades de registro.

As categorias de contexto deste trabalho foram selecionadas a partir da Revisão de Literatura e orientadas igualmente pelas estratégias de busca sendo possível visualizá-las no Quadro 14.

Quadro 14 - Categorias de contexto

Categoria de Contexto	Principais Conceito	Lócus da coleta
Letramento informacional de dados	Quadros conceituais de letramento com características gerais	<i>Frameworks</i> em letramento de dados: (CARLSON <i>et al.</i> , 2011; CALZADA PRADO; MARZAL, 2013; CARLSON; JOHNSTON, 2015; JEFFRYES; JOHNSTON, 2013; FOSTER, 2015; VITAE, 2011;),
	Ambientes Virtuais de Aprendizagem	Recursos de Aprendizagem <i>online</i> : (UNIVERSITY OF OXFORD, 2018-2019); (CONSORTIUM OF EUROPEAN SOCIAL SCIENCE DATA ARCHIVES, 2018), (JOHNSTON; JEFFRYES, 2012); (DATAONE, 2012); (DIGITAL CURATION CENTER, 2014-2019); (RESEARCH DATA NETHERLANDS, <i>online</i>); (FOSTER, 2015), (CORTI <i>et al.</i> , 2014). (UNIVERSITY OF EDINBURGH, 2017); (NECDMC, 2012); (CURATING RESEARCH ASSETS AND DATA USING LIFECYCLE EDUCATION (CRADLE); UNIVERSITY OF NORTH CAROLINA; UNIVERSITY OF EDINBURGH, 2017). Quadro 4
Reúso de dados	Comportamento dos Pesquisadores	Dados secundários sobre prática do reúso. (KIM, 2016[<i>dataset</i>]); (KIM, 2017[<i>dataset</i>]); (BERGHMANS <i>et al.</i> , 2017[<i>dataset</i>]); (LEIDEN UNIVERSITY; ELSEVIER, 2017).

Fonte: Autoria própria (2018).

Sequencialmente apresentam-se os procedimentos para a análise das categorias de contexto indicadas no Quadro 14. A categoria Letramento informacional de dados foi analisada separada em duas partes: a primeira parte da análise se deu com os quadros conceituais de letramento (*frameworks*), descrita na seção 5.9. A categoria de contexto Reúso de dados foi construída com a análise dos conjuntos de dados (*datasets*), descrita na seção 5.10. A segunda parte se deu com a análise do Ambientes Virtuais de Aprendizagem, descrita na seção 5.11.

5.9 PROCEDIMENTOS DE ANÁLISE DE *FRAMEWORKS* DE LETRAMENTO EM DADOS: requisitos de reúso

A Codificação é a primeira etapa, em que o texto é transformado em dados brutos de acordo com a técnica selecionada. A seleção do material foi realizada na etapa do levantamento bibliográfico e na revisão da literatura, a partir da leitura integral dos materiais que permitiu identificar os *frameworks* para letramento em dados, como apresentado no Quadro 14. A partir desse material foi possível identificar os principais elementos para a análise de conteúdo no contexto de reúso de dados.

O *software* NVivo dispõe de uma série de recursos para exploração do material, como frequência de palavras, gráfico de hierarquia, análise de conjuntos, gráficos de relacionamentos, entre outros. Na fase de análise preliminar foram utilizadas as funcionalidades de contagem de palavras do material selecionado, além da pesquisa de palavras nos textos, visualizando as suas relações. No entanto procedeu-se igualmente o processo de forma manual, para efeito comparativo.

A partir da análise quantitativa é possível o estabelecimento das categorias de análise e das unidades de registro, a partir do processamento das palavras mais frequentes por meio de relatórios de recorrência de palavras, disponível no NVivo, e executados nos textos dos *frameworks* analisados. As unidades de contextos foram verificadas no conteúdo dos parágrafos em que a palavra é recuperada, evidenciando a contextualização do termo. As 50 palavras mais recorrentes nos *frameworks* analisados são apresentadas na Figura 19.

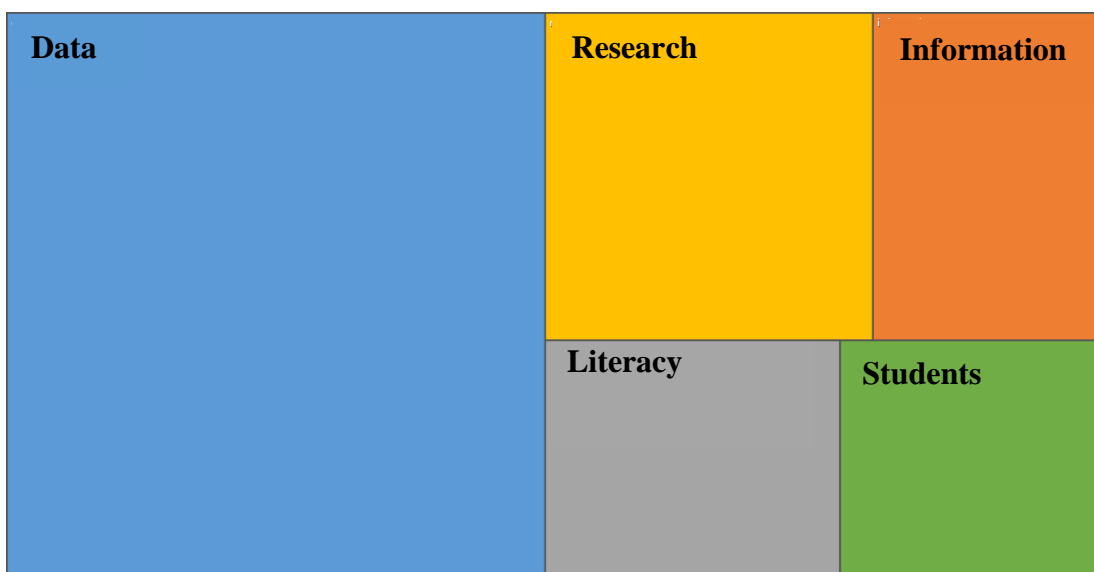
Figura 19 - As 50 palavras mais recorrentes



Fonte: Autoria própria a partir de análises do NVivo (2018).

As categorias de palavras identificadas nos *frameworks* 01 a 06 (CARLSON *et al.*, 2011; CALZADA PRADO; MARZAL, 2013; CARLSON; JOHNSTON, 2015; JEFFRYES; JOHNSTON, 2013; FOSTER, 2015; VITAE, 2011) são apresentadas a seguir. O tamanho de cada quadrante da Figura 20 é estabelecido de acordo com a representatividade de cada palavra, ou seja, os maiores representam os termos mais recorrentes e os menores representam os termos menos recorrentes no conjunto de documentos analisados. Foram identificadas as seguintes palavras mais recorrentes: *data*, (dados) *research* (pesquisa), *information* (informação), *literacy* (letramento) e *students* (estudantes), apresentadas na Figura 20. Essas composições revelam o conteúdo principal tratado no conjunto de documentos: **letramento informacional em dados de pesquisa para estudantes** e coincidem com as categorias identificadas manualmente, representando as categorias de contexto.

Figura 20 - Categorias criadas por palavras mais recorrentes



Fonte: Autoria própria a partir de análises do NVivo (2018).

Ao desdobrar as variações de palavras mais frequentes, uma nova análise por frequência foi realizada, como mostra a Figura 21, resultando em *access data sources*, *data curation*, *data formats*, *using data*, dentre outros exemplos, **representando as categorias de análise**, como retomados mais à frente no Quadro 15.

Figura 21 - Exemplo de variações de palavras mais recorrentes - data

Data	data conver...	data lifec...	rese...	incor...	inclu...	furt...	data...	data ...	data...	sci...	pre...	
	standard ...	data coll...	geoinfor...									da...
using data	source data...	access da...	discipli...									
			data sets	data ...	data...							
data manag...	data formats	external ...	data pr...	beco...	data...	data ...						
	synthesize ...	data vis...	data m...	basic ...	data ...	data ...						
science data li...	specific data	data sha...	standard...	toget...	data ...	data ...						
science data li...	research d...	data rep...	publish...	supp...	data ...	data ...						
			formal d...	relat...	data...	data...						
original resear...	managing ...	core data...	demonst...	origin...	core ...	data ...						
			handli...	bring...	data...	data...						
data preserva...	data produ...	scientific...	data co...	differ...	appr...	data...						

Fonte: Autoria própria a partir de análises do NVivo (2018)

Os desdobramentos de um dos termos indicados na Figura 22 são apresentados no Quadro 15. Neste exemplo, usando a palavra *data* (uma das codificadas pela função automática no *Software NVivo*). São apresentadas as indicações de uso deste termo no texto analisado, bem como a quantidade de referências dos mesmos, na segunda coluna. Para cada um dos termos codificados automaticamente (*data*, *research*, *information*, *literacy* e *students*) o sistema gerou uma lista de palavras decorrentes, a maior recorrência permite a formação de uma categoria.

Quadro 15 - Amostra de palavras mais recorrentes na análise por grupo de frameworks

(continua)

Códigos	Número de referências de codificação
Nós\Frameworks 01-06 Sentenças\data\access data sources	10
Nós\Frameworks 01-06 Sentenças\data\data conversion	12
Nós\Frameworks 01-06 Sentenças\data\data curation	23
Nós\Frameworks 01-06 Sentenças\data\data description	8
Nós\Frameworks 01-06 Sentenças\data\data formats	11
Nós\Frameworks 01-06 Sentenças\data\data librarians	8
Nós\Frameworks 01-06 Sentenças\data\data repository policies	9
Nós\Frameworks 01-06 Sentenças\data\data sharing	9
Nós\Frameworks 01-06 Sentenças\data\data specialists	8
Nós\Frameworks 01-06 Sentenças\data\data visualization	9

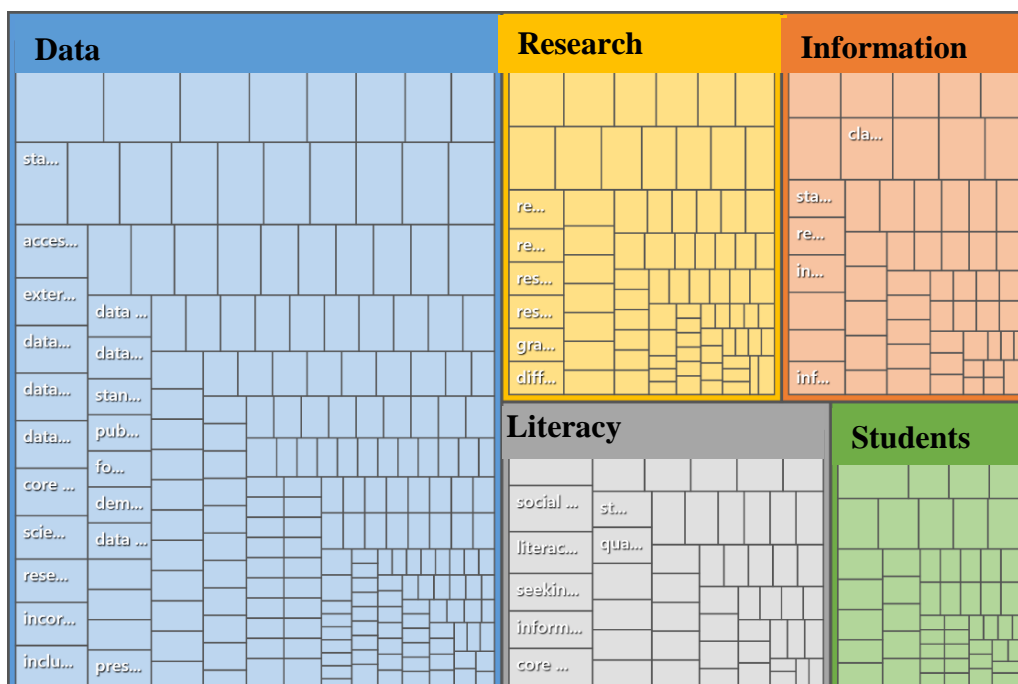
(conclusão)

Códigos	Número de referências de codificação
Nós\Frameworks 01-06 Sentenças\data\disciplinary data repositories	7
Nós\Frameworks 01-06 Sentenças\data\including data citation	8
Nós\Frameworks 01-06 Sentenças\data\managing data	10
Nós\Frameworks 01-06 Sentenças\data\original research data	13
Nós\Frameworks 01-06 Sentenças\data\research data curation	10
Nós\Frameworks 01-06 Sentenças\data\science data life cycle	15
Nós\Frameworks 01-06 Sentenças\data\standard data formats	11
Nós\Frameworks 01-06 Sentenças\data\synthesize data	10
Nós\Frameworks 01-06 Sentenças\data\using data	20

Fonte: Autoria própria a partir de análises do NVivo (2019).

A composição de recorrência de palavras pode ser verificada para cada categoria gerada na análise de termos mais frequentes. A título de exemplo ilustra-se na Figura 22 como cada categoria foi desdobrada em variações dos termos, **resultando nas unidades de registro**, a partir dos termos mais recorrentes por *framework*: *data* (dados - quadrante azul), *research* (pesquisa - quadrante amarelo), *information* (informação - quadrante laranja), *literacy* (letramento - quadrante cinza) e *students* (estudantes - quadrante verde).

Figura 22 - Palavras mais recorrentes – categorias de palavras mais recorrentes



Fonte: Autoria própria a partir de análises do NVivo (2019).

Os termos mais recorrentes por *framework* são registrados na Tabela 13 a seguir, indicados por ordem de recorrência: *data* (dados, 362 recorrências), *research* (pesquisa, 112 recorrências), *information* (informação, 80 recorrências), *students* (estudantes, 52 recorrências), *literacy* (letramento, 49 recorrências), *learning* (aprendizado, 25 recorrências), *statistical* (estatística, 23 recorrências), *skills/abilities* (habilidades, 20 recorrências), *knowledge* (conhecimento, 10 recorrências), *management* (gestão, 10 recorrências) e *impact* (impacto, 8 recorrências).

Tabela 13 - Lista de palavras mais recorrentes na análise por *frameworks* individuais

Termo	Total Recorrências	Framework 01	Framework 02	Framework 03	Framework 04	Framework 05	Framework 06
<i>Data</i>	362	142	115	3	9	39	54
<i>Research</i>	112	35		14	34	19	10
<i>Information</i>	80	38	28		8		6
<i>Students</i>	52	42					10
<i>Literacy</i>	49		44		5		
<i>Learning</i>	25			10			15
<i>Statistical</i>	23		23				
<i>Skills/abilities</i>	20				9	11	
<i>Knowledge</i>	10				10		
<i>Management</i>	10				10		
<i>Impact</i>	8			8			

Fonte: Autoria própria a partir de análises do NVivo (2019).

Após a codificação automática por documentos agrupados (*Frameworks* 01 a 06), se deu a codificação automática dos documentos individualmente. Essa abordagem permite conhecer os termos mais recorrentes dentro de um arquivo sem a interferência dos demais, pois a codificação automática considera a contagem de palavras do conjunto de dados.

A codificação automática consiste na parametrização do *software* NVivo para que, a partir dos seus algoritmos internos, seja feita uma varredura das palavras do texto, sendo agrupadas automaticamente com base na repetição das mesmas e no contexto em que estão empregadas. Esse processo foi realizado nos arquivos dos *Frameworks* 01 ao 06, identificando as palavras mais recorrentes no conjunto desses documentos e sendo gerado para cada classe um nó, ou seja, uma categoria para classificação.

Após a codificação automática, foi realizada igualmente a codificação manual, ou seja, a análise de conteúdo de cada documento e a classificação dos termos localizados, atribuídos manualmente para identificar as unidades de registro. Essa etapa consiste na leitura completa

dos documentos, a fim de que sejam atribuídos “nós” ou características a determinados extratos ou palavras do texto, que configurem elementos de classificação desse conteúdo selecionado.

A metodologia de codificação seguiu as recomendações de Bardin (1995), sendo identificadas as necessidades de letramento e as habilidades requeridas para cada demanda. Os trechos e/ou as palavras nos textos que indicavam as necessidades de letramento foram classificadas manualmente como “nó” Unidades de Registro. Os trechos e/ou as palavras nos textos que indicavam as habilidades requeridas, que normalmente são acompanhadas de exemplos e explicações detalhadas, foram classificadas como o “nó” Unidades de Contexto. As extrações classificadas no “nó” Unidades de Registro somaram 154 referências, integralmente disponíveis no Quadro 40 do Apêndice H. Exemplos das Unidades de Registro, classificadas por cores indicando cada *framework* de onde foi extraída (ver legenda) são apresentados no Quadro 16.

Quadro 16 - Exemplos de unidades de registro de frameworks

UNIDADES DE REGISTRO	FRAMEWORK	UNIDADES DE REGISTRO	FRAMEWORK
<i>Academic literacy and numeracy</i>	F04	<i>Collaboration</i>	F04
<i>Access</i>	F02	<i>Collecting</i>	F02
<i>Access</i>	F02	<i>Collecting</i>	F01
<i>Acquisition</i>	F02	<i>Completing your DMP</i>	F03
<i>Acquisition</i>	F01	<i>Conversion</i>	F01
<i>Acquisition of Data</i>	F01	<i>Critical thinking</i>	F04
<i>Analysing</i>	F04	<i>Critically assess</i>	F02
<i>Analysis</i>	F02	<i>Cultures of Practice</i>	F01
<i>Analysis</i>	F01	<i>Curation</i>	F01
<i>Apply your licence</i>	F03	<i>Customize</i>	F02
<i>Archiving practices</i>	F02	<i>Data Access</i>	F06
<i>Argument construction</i>	F04	<i>Data Analysis</i>	F01
<i>Attribution and co-authorship</i>	F04	<i>Data and users</i>	F02
<i>Authors & Publisher rights</i>	F03	<i>Data collections</i>	F02
<i>Backup</i>	F02	<i>Data Conversion</i>	F01
<i>Best practices for data retention</i>	F01	<i>Data curation</i>	F02

Fonte: Autoria própria a partir de análises do NVivo (2019).

Legenda:

F01	F02	F03	F04	F06
<i>Framework 01</i>	<i>Framework 02</i>	<i>Framework 03</i>	<i>Framework 04</i>	<i>Framework 06</i>

As relações entre os termos são verificáveis em cada documento, sendo possível a extração das unidades de contexto. A seguir estão as extrações classificadas no “nó” Unidades de Contexto, que totalizaram 195 referências. Alguns exemplos de unidades de contexto dos documentos analisados são apresentados no Quadro 17. A relação completa das unidades de contextos extraídas dos *frameworks* está disponível no Quadro 41 do Apêndice I.

Quadro 17 - Exemplo de unidades de contextos de *frameworks*

UNIDADES DE CONTEXTO	FRAMEWORK
<i>Ability to access data sources appropriate to the information needed</i>	F02
<i>Ability to apply results to learning, decision making or problem-solving</i>	F02
<i>Ability to critically assess data and their sources</i>	F02
<i>Ability to determine and use suitable research methods</i>	F02
<i>Ability to determine when data are needed</i>	F02
<i>Ability to handle and analyze data</i>	F02
<i>Ability to identify the context in which data are produced and reused (data lifecycle</i>	F02
<i>Ability to plan, organize and self-assess throughout the process</i>	F02
<i>Ability to present quantitative information (specific data, tables, graphs, in reports and similar)</i>	F02
<i>Ability to recognize source data value, types and formats</i>	F02
<i>Analyse existing OS Guidelines</i>	F03
<i>Analyses policies and understands the wider contexts in which they are situated.</i>	F04
<i>Apply lessons learnt for your discipline/project</i>	F03
<i>Apply licence and apply to your content</i>	F03
<i>Apply the RDM Plan to your discipline/project</i>	F03

Fonte: Adaptado dos *frameworks* de Calzada Prado e Marzal (2013), Carlson *et al.* (2011), Carlson e Johnston (2015), Foster (2015), Jeffreyes e Johnston (2013) e Vitae (2011), a partir de análises do NVivo (2019).

A classificação das unidades de registro foi a base para identificar as Categorias de Análise, em que uma camada mais ampla do que de unidades de registro é constituída, de acordo com a metodologia de Bardin (1995). A identificação das Categorias de Análise partiu do agrupamento semântico e léxico da terminologia usada nas Unidades de Registro, recuperadas da codificação manual dos *frameworks*.

As Categorias de Análise a partir do conjunto de *frameworks* analisados são apresentadas no Quadro 18: *Access, Acquisition, Analysis, Archiving*, com alguns exemplos de unidades de registro, o agrupamento de todas as Unidades de Registros em Categorias de Análises está integralmente disponível no Quadro 42 do Apêndice J, sendo: *Access, Acquisition, Analysis, Archiving, Curation, Data management, Documentation, Licences, Metadata, Open Science, Policy, Quality, Reuse, Sharing, Storage, Using, Visualization*.

Quadro 18 - Exemplo de agrupamento de categorias de análise de frameworks

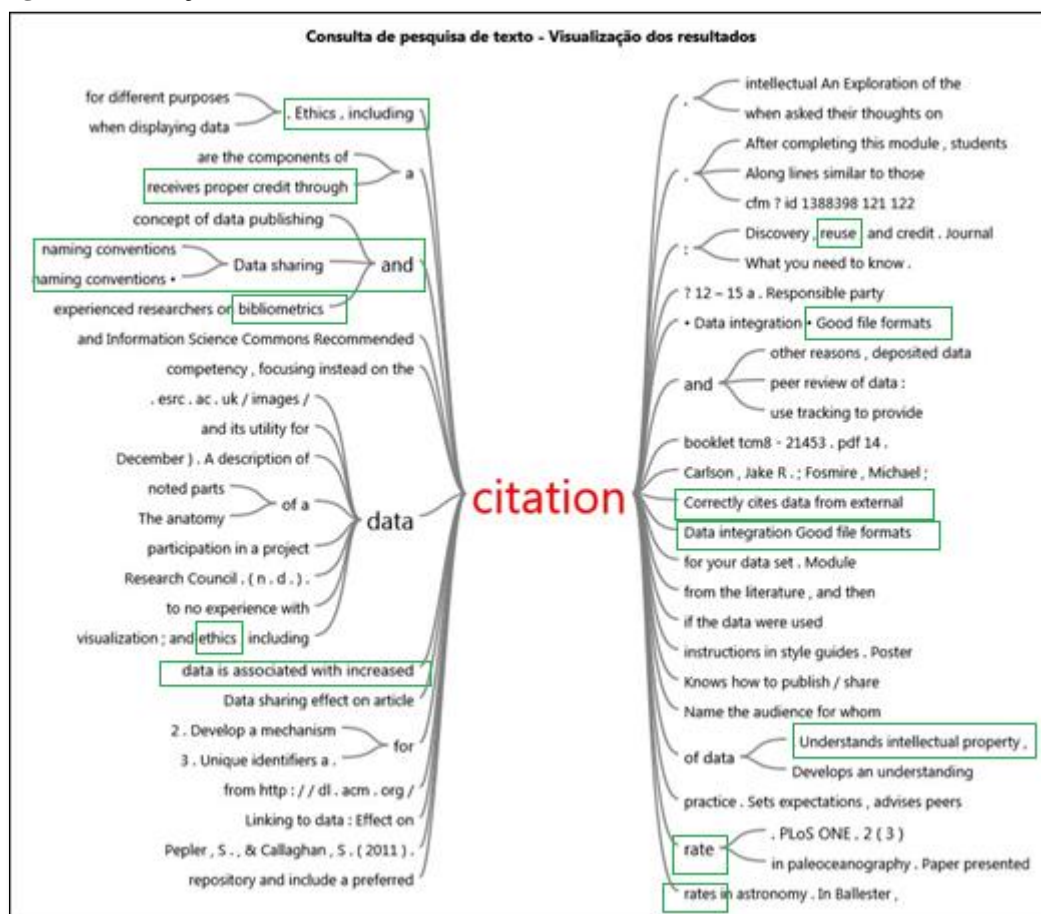
CATEGORIAS DE ANÁLISE	UNIDADES DE REGISTRO	CATEGORIAS DE ANÁLISE	UNIDADES DE REGISTRO
Access	Access	Analysis	Analysing
	Access		Analysis
	Critically assess		Analysis
	Data Access		Argument construction
	Discovery		Critical thinking
	Discovery		Data Analysis
	Information seeking		Evaluate data
	Locate		Evaluating
Acquisition	Acquisition		Evaluating
	Acquisition		Evaluating
	Acquisition of Data		Evaluation
	Collecting		Interpret
	Collecting		Interpretation
	Extract		Problem solving
	Finding and/or obtaining data	Reading, interpreting and evaluating data	

Fonte: Autoria Própria a partir de análises do NVivo (2019).

Uma versão completa que também apresenta as unidades de registro e as unidades de contexto extraídas da codificação manual a partir da leitura completa e da combinação do material utilizado está no Quadro 43, disponível no Apêndice K.

Ao pesquisar as palavras mais selecionadas, é apresentada uma relação semântica deste termo no contexto em que aparece, possibilitando inferências sobre os tipos de relações às quais o termo está subordinado, como por exemplo o termo *citation*, na Figura 23.

Figura 23 - Relações semânticas dos termos selecionados- *citation*



Fonte: Autoria própria a partir de análises do NVivo, (2019).

A análise das relações envolvendo o termo *citation* permite inferir, por exemplo, que um ou mais documentos analisados tratam *citation* no aspecto da citação dados (a partir da forte relação com a palavra *data*, em destaque). Também que citação está relacionada com compartilhamento de dados e convenções de nomenclaturas (*naming conventions*), formato de arquivos (*file formats*) e reuso de dados (*reuse*). Relaciona-se com questões de ética e propriedade intelectual (*Ethics, Intellectual property*) e como critérios de ranqueamento de publicações (*bibliometrics, rate*). A partir desta metodologia apresentada é realizada a análise de conteúdo dos modelos de letramento dos *frameworks* (OE2). Lembra-se que a coleta dos *frameworks* se deu a partir do resultado da revisão sistemática da literatura, em que foi possível identificar as propostas dos quadros conceituais de letramento informacional em dados.

5.9.1 Análises dos *Frameworks* sobre Letramento em Dados

Após a análise e a extração dos conceitos combinados dos *frameworks* estudados, o

resultado das categorias de contexto *framework* é demonstrando sequencialmente, incluindo as categorias de análise e as unidades de registro identificadas por meio da análise de conteúdo, sendo:

- a) **Categorias de análise:** *Access* (Acesso), *Acquisition* (Aquisição), *Analysis* (Análise), *Archiving* (Arquivamento), *Curation* (Curadoria), *Data management* (Gerenciamento de Dados), *Documentation* (Documentação), *Licences* (Licenças), *Metadata* (Metadados), *Open Science* (Ciência Aberta), *Policy* (Política), *Quality* (Qualidade), *Reuse* (Reutilização), *Sharing* (Compartilhamento), *Storage* (Armazenamento), *Using* (Uso), *Visualization* (Visualização).
- b) **Unidades de registro:** *Access data sources* (Acessar fontes de dados), *Discovery* (Descoberta), *Access data* (Acessar dados), *Reading, Interpreting/ Understanding data* (Ler, Interpretar / Entender dados), *Evaluating* (Avaliar), *Critical thinking* (Pensamento crítico), *Obtaining data* (Obter dados), *Collecting Data* (Coletar dados), *Plan for archiving* (Planejar arquivamento), *Archiving practices* (Práticas de arquivamento), *Data Preservation*, (Preservação de dados), *Techniques* (Técnicas), *Conversion* (Conversão), *Formats* (Formatos), *Merging* (Mesclar), *Processing* (Processar), *Synthesize* (Sintetizar), *Data Management Plan* (Plano de gerenciamento de dados), *Directory structures* (estruturas de diretório), *Naming conventions* (convenções de nomenclatura), *Organizational planning* (planejamento organizacional), *Standards* (padrões), *Data sets description* (descrição de conjuntos de dados), *Documentation Methods* (métodos de documentação), *Apply your licence* (aplicar sua licença), *Citation* (citação), *Intellectual Property Rights* (direitos de propriedade intelectual), *Licenses types and copyright* (tipos de licenças e direitos autorais), *Publisher rights* (direitos do editor), *Using data ethically* (uso de dados eticamente), *Metadata Description* (descrição de metadados), *Metadata Standards* (padrões de metadados), *Concept of Open Science* (Conceito de Ciência Aberta), *Open Access* (Acesso Aberto), *Open Data Journals* (Revistas de Dados Abertos), *Open Data Standards* (Padrões de Dados Abertos), *Open Data use and reuse* (Uso e reutilização de Dados Abertos), *Funding policies* (Políticas de financiamento), *Quality Assurance* (Garantia de Qualidade), *Research methods – practical application* (Métodos de pesquisa - aplicação prática), *Research methods – theoretical knowledge* (Métodos de pesquisa - conhecimento teórico), *Reproducible Research* (Pesquisa reproduzível), *Legal requirements* (Requisitos legais), *Reuse policies* (Políticas de reutilização), *re-sharing* (re-compartilhamento), *Backup* (backup), *Security* (segurança),

Databases (bancos de dados), *Data collections* (coletas de dados), *Visualization tools* (ferramentas de visualização).

Optou-se por manter a descrição em inglês, pois a análise léxica poderia ser alterada na tradução. O relacionamento entre as Categorias de Análise e as Unidades de Registro é apresentado no Quadro 42. A partir das categorias de análise e das unidades de registro apresentada no Quadro 41, pode-se indicar as unidades de registro que se relacionam com o reúso de dados. O Quadro 19 apresenta os requisitos para o reúso de dados a partir do levantamento nos quadros conceituais (*frameworks*) de letramento informacional em dados, conforme o Objetivo Específico 2 (OE2). O Quadro 19 apresenta os requisitos para o reúso de dados a partir do levantamento nos quadros conceituais (*frameworks*) de letramento informacional em dados, conforme o Objetivo Específico 2 (OE2).

Quadro 19 - OE2: Requisitos para o reúso de dados – análise dos quadros conceituais (*frameworks*) combinados

Categorias de Análise	Unidades de Registro	Aplicável ao reúso de dados
1. Access	1.1 Access data sources	Sim
	Discovery	Sim
	Access data	Sim
2. Analysis	Reading	Sim
	Interpreting/ Understanding data	Sim
	Evaluating	Sim
	Critical thinking	Sim
3. Acquisition	Obtaining data	Sim
	Collecting Data	Sim
4. Archiving	Plan for archiving	Não
	Archiving practices	Não
	Data Preservation Techniques	Não
5. Curation	Conversion	Sim
	Formats	Sim
	Merging	Sim
	Processing	Sim
	Synthesize	Sim
6. Data Management	Data Management Pela	Sim
	Directory structures	Não
	Naming conventions	Sim
	Organizational planning	Não
	Standards	Não
7. Documentation	Data sets description	Sim
	Documentation Methods	Não
8. Licences	Apply your licence	Não
	Citation	Sim
	Intellectual Property Rights	Sim
	Licenses types and copyright	Sim
	Publisher rights	Sim
	Using data ethically	Sim
9. Metadata	Metadata Description	Não
	Metadata Standards	Sim
10. Open Science	Concept of Open Science	Sim
	Open Access	Sim
	Open Data Journals	Sim
	Open Data Standards	Não
	Open Data use and reuse	Sim
11. Policy	Funding policies	Não
12. Quality	Quality Assurance	Sim
	Research methods – practical application	Sim
	Research methods – theoretical knowledge	Sim
13. Reuse	Reproducible Research	Sim
	Legal requirements	Sim
	Re-use policies	Sim
14. Sharing	Re-sharing	Sim
15. Storage	Backup	Não
	Security	Não
16. Using	Databases	Sim
	Data collections	Sim
17. Visualization	Visualization tools	Sim

Fonte: Adaptado dos *frameworks* de Calzada Prado e Marzal (2013), Carlson *et al.* (2011), Carlson e Johnston (2015), Foster (2015), Jeffreyes e Johnston (2013) e Vitae (2011), a partir de análises do NVivo (2018).

Em seguida descreve-se o procedimento para análise dos dados secundários (*datasets*) cuja finalidade é identificar as perspectivas dos pesquisadores sobre o reúso de dados (OE3).

5.10 PROCEDIMENTOS DE ANÁLISES DE DADOS SECUNDÁRIOS – CONJUNTOS DE DADOS

A análise dos *datasets* seguiu a sequência identificada previamente como resultado da análise dos *frameworks*, na seção 5.9, observando cada categoria de análise e as suas respectivas unidades de registro. Cada um dos requisitos (dentre os que foram identificados na análise dos *frameworks*), foram aplicados na etapa de análises dos dados secundários realizada pela pesquisadora. O Quadro 20 apresenta essa sequência e as medidas adotadas para a execução do reúso de dados nesta pesquisa.

Quadro 20 - Requisitos aplicados nesta pesquisa para a análise dos dados secundários

(continua)

Categorias de Análise	Unidades de Registro	Como os requisitos foram aplicados ao reúso de dado nesta pesquisa
1. <i>Access</i>	<i>Access data sources</i>	Foram identificados repositórios com cobertura da área de ciências sociais no indexador www.r3data.org . Em seguida foi realizada uma pesquisa no repositório de Ciências Sociais Inter-university Consortium for Political and Social Research (ICPSR)
	<i>Discovery</i>	Foram localizados três conjuntos de dados de pesquisas sobre compartilhamento e reúso de dados (KIM, 2016 [<i>datasets</i>]) (KIM, 2017 [<i>datasets</i>]); (BERGHMANS <i>et al.</i> , 2017 [<i>datasets</i>])
	<i>Access data</i>	Os dados foram acessados via cadastro no portal ICPSR
2. <i>Analysis</i>	<i>Reading</i>	Foi feita a leitura dos artigos relacionados aos <i>datasets</i> , além da leitura do conteúdo dos dados e da documentação disponível
	<i>Interpreting/ Understanding data</i>	Foi realizada a busca de dados de pesquisa sobre comportamentos dos pesquisadores em relação aos reúso de dados, portanto, necessária a análise e o entendimento se este era o tema do trabalho e dos dados localizados. Para tal se observou as perguntas do instrumento de pesquisa disponível dentre a documentação dos dados
	<i>Evaluating</i>	Buscou-se evidenciar a perspectiva do pesquisador em relação as práticas com os dados de pesquisa, além da confirmação da familiaridade do público pesquisado nos questionários, se esta público era consciente sobre o tema de reúso de dados
	<i>Critical thinking</i>	Foram avaliadas as metodologias utilizadas e descritas em cada artigo relacionado com os dados, identificando o universo de pesquisa e a validade da amostra, sua representatividade quantitativa e qualitativa para o reúso pretendido
3. <i>Acquisition</i>	<i>Obtaining data</i>	Esses requisitos foram executados de forma simultânea, na medida em que os arquivos de dados eram obtidos por meio de <i>downloads</i> , e em seguida armazenados na pasta do <i>drive</i> virtual da pesquisadora
	<i>Collecting Data</i>	
4. <i>Archiving</i>	<i>Plan for archiving</i>	Os dados foram armazenados em nuvem e no disco rígido externo, além do computador pessoal. Uma forma de facilitar a atualização dos arquivos é o uso de <i>drivers</i> em nuvem, sincronizados com os arquivos salvos na máquina. Neste estudo foi utilizado o aplicativo <i>Backup and Sync</i> , <i>driver</i> de sincronização dos arquivos no computador com o Google <i>cloud</i> para contas de pessoas físicas.
	<i>Archiving practices</i>	
	<i>Data Preservation Techniques</i>	Os dados foram convertidos dos formatos proprietários: planilhas de dados em Microsoft Excel e documentos do Microsoft Word.

(continua)

Categorias de Análise	Unidades de Registro	Como os requisitos foram aplicados ao reuso de dado nesta pesquisa
5. <i>Curation</i>	<i>Conversion</i>	As extensões dos arquivos foram analisadas e a conversão para formatos abertos (OpenOffice). Foram geradas cópias para os arquivos que estavam em formato proprietários Microsoft Excel e Microsoft Word
	<i>Formats</i>	
	<i>Merging</i>	Não houve fusão de arquivos
	<i>Processing</i>	Posteriormente, os conteúdos dos dados em formato Microsoft Excel .csv (separado por vírgula) foram processados e transformados em colunas para que as sínteses e as novas análises fossem viáveis, aplicando-se os recursos de filtros e contagem automática das colunas selecionadas
	<i>Synthesize</i>	
6. <i>Data Management</i>	<i>Data Management</i>	Neste requisito previu-se a atualização do Plano de Gestão de Dados com a informação dos <i>datasets</i> localizados
	<i>Pela</i>	Os arquivos foram estruturados em pastas dentro do <i>drive</i> da elaboração da Tese.
	<i>Directory structures</i>	
	<i>Naming conventions</i>	Os arquivos foram renomeados de acordo com padrões de nomenclatura que identificassem o conteúdo e a autoria
	<i>Organizational planning</i>	Não aplicável neste estudo
<i>Standards</i>	Foram usados padrões de identificação de conteúdo e autoria (AUTOR, ano.*)	
7. <i>Documentation</i>	<i>Data sets description</i>	Foi preciso verificar no questionário e no arquivo de dados, quais eram os códigos dos campos e a descrição correspondente para as perguntas previamente selecionadas para reuso
	<i>Documentation Methods</i>	Neste requisito o foi feita uma análise dos instrumentos de coleta de dados (questionários) que foram usados para elaboração das perguntas, a fim de selecionar as questões que mais teriam aderência ao estudo pretendido. Outros documentos auxiliares foram importantes para identificar, por exemplo, a codificação de respostas e não apenas de perguntas, como os números correspondentes às áreas do conhecimento a qual cada entrevistado pertencia
8. <i>Licences</i>	<i>Apply your licence</i>	Não aplicável neste estudo, mas foi necessário identificar qual a licença selecionada pelos autores dos dados para reusar em conformidade
	<i>Citation</i>	Neste requisito foi possível verificar o tipo de licença permitida para reuso, assim como a forma de citação dos conjuntos de dados
	<i>Intellectual Property Rights</i>	Importante notar que nem sempre a Propriedade intelectual do documento primário que apresenta os resultados é a mesma do conjunto de dados à este vinculado (<i>Intellectual Property Rights</i>). Por exemplo, a análise e publicação do conjunto de dados BERGHMANS <i>et al.</i> (2017) [<i>datasets</i>] está publicada sob a propriedade intelectual de outra entidade, no caso o relatório em nome da Leiden University e Elsevier (2017)
	<i>Licenses types and copyright</i>	
	<i>Publisher rights</i>	Não aplicável neste estudo
	<i>Using data ethically</i>	Um ponto importante foi identificar se havia o anonimato necessário nas respostas, que evitassem a identificação dos participantes. Os autores dos conjuntos de dados haviam identificados os respondentes por códigos, o que facilitou esse cuidado
9. <i>Metadata</i>	<i>Metadata Description</i>	Este requisito ajudou a compreender os elementos de identificação do teor dos conjuntos de dados, mas não que se precisasse saber antecipadamente qual o padrão de campos e se o preenchimento estava completo
	<i>Metadata Standards</i>	

(conclusão)

Categorias de Análise	Unidades de Registro	Como os requisitos foram aplicados ao reúso de dado nesta pesquisa
10. <i>Open Science</i>	<i>Concept of Open Science</i>	Esses requisitos são elementos mais conceituais que envolvem a disponibilidade de dados, porém, podem ser percebidos por todo o processo ao se localizar e ter a disponibilidade de reúso dos conjuntos de dados
	<i>Open Access</i>	
	<i>Open Data Journals</i>	
	<i>Open Data Standards</i>	
	<i>Open Data use and reuse</i>	
11. <i>Policy</i>	<i>Funding policies</i>	Foi importante compreender as políticas de financiamento, entender que os dois conjuntos de dados utilizados estavam disponíveis porque atendiam à essas exigências de publicidade e de abertura de dados vinculados às pesquisas que foram financiadas com recursos públicos e estavam sujeitas à essas contrapartidas
12. <i>Quality</i>	<i>Quality Assurance</i>	Considerou-se as instituições às quais as pesquisas e os dados estão vinculados, sendo a Syracuse University e Leiden University como critérios de qualidade para seguir com o reúso
	<i>Research methods – practical application</i>	Identificou-se a consistência metodológica das pesquisas utilizadas e da documentação de apoio aos questionários
	<i>Research methods – theoretical knowledge</i>	Foi identificada a base teórica da construção das perguntas do instrumento de coleta de dados, reproduzido no Anexo C .
13. <i>Reuse</i>	<i>Reproducible Research</i>	Esse requisito baseou-se nas condições de reúso da pesquisa, que busca identificar comportamentos e percepções similares dos pesquisadores em coleta de dados previamente realizada
	<i>Legal requirements</i>	Não houve qualquer impedimento para o aproveitamento das respostas, tanto por parte da instituição em que a pesquisa foi realizada, como da instituição vinculada ao estudo para o qual os dados seriam reusados.
	<i>Re-use policies</i>	
14. <i>Sharing</i>	<i>Re-sharing</i>	O Plano de gestão de dados deste trabalho prevê que os resultados desta pesquisa sejam compartilhados
15. <i>Storage</i>	<i>Backup</i>	Ainda que este requisito não se relacione ao reúso diretamente, será aplicado aos novos dados na fase de compartilhamento, sendo necessárias medidas para o armazenamento dos dados. As condições estão listadas no Plano de gestão de dados deste trabalho
	<i>Security</i>	
16. <i>Using</i>	<i>Databases</i>	Este requisito considerou o uso adequando de bases de dados e coleções de dados, que foi demonstrada nesse trabalho, iniciando pela busca de repositório disciplinar na área das Ciências Sociais
	<i>Data collections</i>	
17. <i>Visualization</i>	<i>Visualization tools</i>	A visualização foi executada na apresentação de gráficos e imagens para a divulgação dos resultados das análises

Fonte: Autoria própria (2019) a partir do reúso de dados dos *datasets* de (KIM, 2016 [*datasets*]) (KIM, 2017 [*datasets*]); (BERGHMANS *et al.*, 2017 [*datasets*]).

Os resultados desses levantamentos foram reanalisados sob o aspecto de quais necessidades de letramento poderiam ser descobertas de acordo com as perspectivas dos pesquisadores, quando estes relatam suas práticas de compartilhamento e de reúso de dados. Esta reanálise foi feita a partir da leitura minuciosa de todas as perguntas dos questionários aplicados aos pesquisadores, indentificando questões que abordassem as competências

necessárias, as dificuldades e as possíveis barreiras encontradas, as habilidades requeridas pelos pesquisadores nesse processo de compartilhamento e de reuso que estivessem incorporados nesses questionários Logo, o processo foi de Análise de Conteúdo, cujas categorias de análise foram especificamente: competências e habilidades, dificuldades e barreiras, compartilhamento e reuso. As unidades de contexto foram, como já mencionado, os questionários de análise de dados Scientists' Data Sharing Behavior Survey (KIM, 2016), Data Reuse Survey Instrument (KIM, 2017) e Research data sharing and reuse (BERGHMANS *et al.*, 2017), disponíveis no Anexo A – Instrumento de coleta de dados 1, Anexo B – Instrumento de coleta de dados 2 e Anexo C – Instrumento de coleta de dados 3.

A análise de dados secundários (*datasets*) permite evidenciar pontos negativos e positivos das percepções dos pesquisadores sobre o tema do reuso de dados em especial para a Categoria de Análise Acesso (*Access*), dentro da unidade de registro descoberta (*Discovery*). As perguntas selecionadas para reuso das respostas disponíveis nos conjuntos de dados foram selecionadas dos instrumentos de coleta de dados 1, de Kim (2016), que está disponível Anexo A, do instrumento de coleta de dados 2, de Kim (2017), que está disponível no Anexo B, do instrumento de coleta de dados 3, de Berghmans *et al.* (2017), [*datasets*];) que está disponível no Anexo C. Os arquivos dos *datasets* estão disponíveis para *downloads* nos *links* das referências, sem quaisquer restrições.

Estes conjuntos de dados foram reusados neste trabalho também para permitir a percepção deste método, executado de forma prática. Ao se analisar estes conjuntos disponíveis, as seguintes etapas foram executadas para todos os *datasets*: seleção de campos pertinentes ao reuso de dados (de acordo com as perguntas dos instrumentos de pesquisa), aplicação de filtros com uso do *Software* Excel com a finalidade de realizar a contagem de frequência de respostas e assim proceder com a análise de conteúdo relacionando as percepções dos pesquisadores com os requisitos de letramento identificados na etapa anterior.

Para a aplicação dos filtros, foi necessário que o conteúdo do arquivo estivesse em formato de colunas e linhas para que as novas análises fossem viáveis. Os arquivos cujos conteúdos estavam em formato com extensão .csv (separado por vírgulas) foram tabulados e transformados em arquivos organizados por colunas. De outra forma não foi possível filtrar as respostas, pois as informações listadas e separadas por vírgulas (formato .csv) não permitiram que as mesmas fossem agrupadas ou ordenadas automaticamente. Foram aplicados os recursos de filtros e a contagem automática das colunas selecionadas, como pode ser visualizado na Figura 24. Para essa transformação, foram necessários os seguintes comandos: 1) visualização do caractere que separa os campos no texto; 2) identificação das colunas que separam os

conteúdos; 3) seleção da coluna de conteúdo; 4) seleção da função para alterar a coluna; 5) seleção de filtros que se deseja verificar nas colunas; 6) seleção de dados para análise marcados nas colunas; 7) aplicação de filtros que se deseja analisar.

Figura 24 - Exemplo de arquivo .CSV (1) transformado em arquivo em colunas (2)

3 - Seleção da coluna em que os dados estão dispostos

4 - Seleção da função para alterar para colunas

2 - colunas separadas por vírgulas, codificadas de acordo com questionário

1 - dados separados por vírgulas

5 - Seleção de colunas por disciplina

6 - Dados em colunas, seleção de respostas para análise

7 - Aplicação de filtros por pergunta a ser analisada

ID	Discipline	MainDiscipline	Usefulness1	Usefulness2	Usefulness3	Concern1	Concern2	Concern3	Effort1	Effort2	Effort3	Climate1	Climate2	Climate3	Reposi1	Reposi2	OrgRes1	OrgRes2	OrgRes3	
1	20	6	2	2	2	2	2	2	2	2	2	2	2	2	1	4	3	4	2	3
2	49	6	1	1	1	1	2	2	2	4	3	5	3	3	2	5	3	4	3	1
3	29	3	1	1	2	4	5	2	3	1	2	2	2	1	5	3	4	3	1	
4	62	2	2	4	1	1	1	1	5	2	4	4	4	5	5	4	4	1	4	3
5	562	2	3	3	3	3	3	3	2	2	2	2	2	2	3	3	1	5	3	4
6	59	2	2	1	1	4	3	2	4	3	2	2	2	1	2	1	3	1	5	3
7	21	2	3	3	4	2	2	2	2	2	2	3	2	2	3	2	2	4	3	4
8	33	2	2	5	3	3	1	1	1	2	2	2	3	4	4	1	1	1	4	3

Fonte: Autoria Própria (2019).

Os filtros selecionados para as respostas do questionário de Kim (2016 [*datasets*]) foram os resultados da soma das opções de respostas que representavam maior concordância com as afirmativas das questões, sendo elas: 7 – concordo totalmente (7 *Strongly Agree*), 6 – concordo moderadamente (6 *Moderately Agree*) e 5 – concordo parcialmente (5 *Slightly Agree*). Dentre as 2.470 respostas válidas, foram filtradas apenas as que tinham como indicação de posição PhD/Doutor (*PhD/Doctoral Degree*) para os respondentes, sendo esta a posição selecionada pelo próprio autor dos dados. A aplicação destes filtros em conjunto resultou em 2.202 respostas válidas para compor a análise. Um sub extrato dessa amostra foi aplicado para a disciplina das Ciências Sociais, totalizando 122 respostas selecionadas com o critério de segmentação para esta disciplina.

As perguntas selecionadas foram as com os seguintes IDs⁴⁵: *Metadata1, Metadata2 e Metadata3; Repository1, Repository2 e Repository3; Altruism1, Altruism2 e Altruism3; Benefit1, Benefit2, Benefit3 e Benefit4; Risk1, Risk2, Risk3 e Risk4; Effort1, Effort2, Effort3 e Effort4; Attitude1, Attitude2, Attitude3 e Attitude4; AltruismCommunity1, AltruismCommunity2, AltruismCommunity3 e AltruismCommunity4; MethodDisciplinaryRepository, MethodInstitutionalRepository, MethodPublicWeb, MethodSupplement, MethodRequested, MethodResponded; Resource1, Resource2, Resource3 e Resource4*. Os resultados da análise do conjunto de dados de Kim (2016 [*datasets*]) estão na Tabela 21.

Os filtros selecionados para as respostas do conjunto de dados de Kim (2017 [*datasets*]), conjunto com 1.240 respostas válidas e utilizadas nesta análise, foram os resultados da soma das opções de maior concordância com as afirmativas das questões, sendo: 1 - concordo totalmente (1 *Strongly Agree*) e 2 – concordo (2 *Agree*). Seguindo a classificação do autor, todas as repostas válidas já estavam categorizadas pela posição de Phd/Doutor (*PhD/Doctoral Degree*). Destes, 176 representavam a área das Ciências Sociais. Os campos selecionados foram: IDs: *Usefulness1, Usefulness2, Usefulness3, Concern1, Concern2, Concer3, Effort1, Effort2, Repository1, Repository2, OrgRoesource1, OrgResource2, OrgResource3, Resource4, Intention1, Intention2 e Intention3*). Os resultados do conjunto de dados de Kim (2017 [*datasets*]) são apresentados na Tabela 22.

Para os dados de Berghmans *et al.* (2017), conjunto de repostas com 1.200 respostas válidas, apenas a opção 1 foi selecionada, visto que esta indicava a escolha de determinada alternativa como resposta pelos entrevistados. As alternativas de respostas deste questionário

⁴⁵ ID refere-se a um código que identifica as perguntas selecionadas para reúso, conforme convencionado nos anexos A, B e C. O uso do ID evita a repetição da pergunta por extenso.

não foram estruturadas em classificação por níveis de concordância/discordância. A quantidade representada pela área das Ciências Sociais nesta pesquisa foi de 75 respondentes. As perguntas selecionadas para esta análise foram os seguintes IDs: Q1a_1 a Q1a_5, Q1b_1 a Q1b_8, Q1c_1 a Q1c_8, Q1ei_1 a Q1ei_99, Q1eii_1 a Q1eii_99, Q1f_1 a Q1f_3, Q2a_1 a Q2a_10, Q2c_1 a Q2c_99, Q2e_1 a Q2e_3, Q3d_1 a Q3d_99, Q3e_1 a Q3e_9 e Q3f_1 a Q3f_4. A Tabela 23 apresenta os resultados do conjunto de dados de Berghmans *et al.* (2017 [*datasets*]).

A partir destes procedimentos apresentados, os resultados de análise dos três *datasets* (OE3) estão descritos na seção 6.1. Sequencialmente apresentam-se os procedimentos para a análise comparada dos Recursos de aprendizagem *online* (OE4), em que foi possível verificar as competências abordadas por cada curso disponível em acesso aberto. Os resultados desta análise estão na seção 6.2.

5.11 PROCEDIMENTOS DAS ANÁLISES DOS RECURSOS DE APRENDIZAGEM ONLINE

A coleta de dados sobre os recursos de aprendizagem *online* se deu a partir da revisão sistemática da literatura, em que foi possível identificar os projetos de letramento em gestão de dados, disponíveis em acesso aberto, que foram listados no Quadro 4. Esta coleta, relembra-se, atende ao objetivo específico (OE4) que trata de verificar se os requisitos identificados nas etapas anteriores estão abordados nos cursos.

Os recursos de aprendizagem *online* foram rotulados da seguinte forma para a etapa da análise:

- Recurso A (R.A): Bodleian Data Library – University of Oxford (UNIVERSITY OF OXFORD, 2018-2019);
- Recurso B (R.B): CESSDA Consortium of European Social Science Data Archives ERIC – European Research Infrastructure Consortium. (CONSORTIUM OF EUROPEAN SOCIAL SCIENCE DATA ARCHIVES, 2018);
- Recurso C (R.C): Data Management Course (JOHNSTON; JEFFRYES, 2012);
- Recurso D (R.D): DataONE Data Observation Network for Earth (DATAONE, 2012);
- Recurso E (R.E): Digital Curation Center - DC 101 materials: Access and Reuse (DIGITAL CURATION CENTER, 2014-2019);
- Recurso F (R.F): Essentials 4 Data Support (RESEARCH DATA NETHERLANDS, *online*);

- Recurso G (R.G):Facilitate Open Science Training for European Research (FOSTER, 2015);
- Recurso H (R.H): Managing and Sharing Research Data handbook (CORTI *et al.*, 2014);
- Recurso I (R.I): MANTRA (UNIVERSITY OF EDINBURGH; EDINA DATA LIBRARY, 2017);
- Recurso J (R.J): New England Collaborative Data Management Curriculum (NECDMC, 2012);
- Recurso K (R.K): Research Data Management and Sharing - Coursera (CURATING RESEARCH ASSETS AND DATA USING LIFECYCLE EDUCATION (CRADLE); UNIVERSITY OF NORTH CAROLINA; UNIVERSITY OF EDINBURGH, 2017).

A partir dos requisitos levantados na etapa de análise dos *frameworks*, seguiu-se a análise pela fase (i) da organização da análise do material disponível nos recursos de aprendizagem *online*. Os arquivos pesquisados foram obtidos no sítio de cada um dos recursos de aprendizagem *online*, sendo agrupados no *software* NVivo em pastas de acordo com as suas hierarquias para proceder com as outras etapas da análise.

A fase (ii), da exploração do material, se deu por meio da funcionalidade de codificação automática do *software* NVivo, que apresentou resultados limitadores para o objetivo específico (OE4) vinculado à esta etapa.

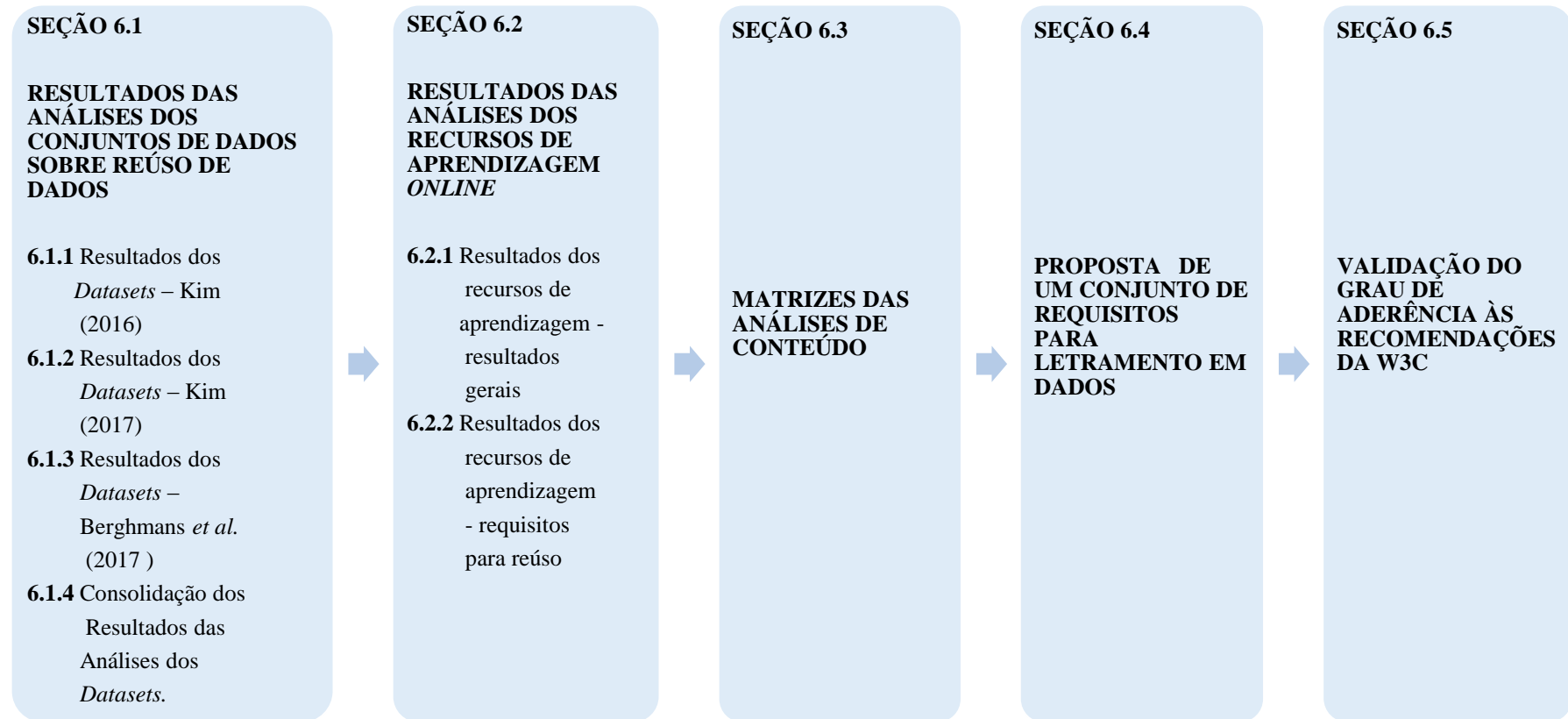
Seguindo com a fase (iii), da administração das técnicas de decisão, em que a classificação em categorias seguiu a estrutura de (a) categorias de contexto (formado pelo tipo de recurso de aprendizado *online*), (b) categorias de análise (formada pelos requisitos) e (c) unidades de registro (formada pelas competências dos requisitos). Desta forma, foram elaboradas estratégias de busca para cada requisito, de acordo com as suas competências, e aplicadas às pesquisas nos recursos de aprendizagem *online*.

As estratégias de busca foram construídas com uso de recursos booleanos, como o operador *OR* (OU) que amplia os resultados pelo uso de sinônimos ou variações de grafia das palavras pesquisadas. As estratégias de busca foram baseadas nas Categorias de Análise: *Access, Analysis, Acquisition, Archiving, Curation, Data Management, Documentation, Licences, Metadata, Open Science, Policy, Quality, Reuse, Sharing, Storage, Using e Visualization*, e as Unidades de Registros identificadas na etapa anterior de análise dos *frameworks*, listadas no Apêndice K.

A busca pelos requisitos identificados previamente, na análise dos *frameworks* (seção 5.9), foi aplicada à pesquisa direcionada ao conteúdo dos documentos dos recursos de aprendizagem *online*. A relação dessas estratégias, desdobrada por requisito e por competência, está disponível no Apêndice O. Os resultados completos dessas análises podem ser vistos na seção 6.2.

Apresentam-se, a seguir, no capítulo 6, os resultados das análises de dados realizadas, atendendo aos objetivos específico OE3, OE4, OE5 e OE6, propostos nesta pesquisa. Orientando a leitura, os resultados das análises dos conjuntos de dados secundários são apresentados na seção 6.1. Os resultados dos recursos de aprendizagem *online* estão listados na seção 6.2. Uma matriz de análise é apresentada na seção 6.3, em que todos os resultados são consolidados. O conjunto de requisitos propostos para o letramento em dados está na seção 6.4. O capítulo é finalizado com a validação dos requisitos segundo as recomendações da W3C, na seção 6.5. A estrutura do capítulo e as subseções seguintes são apresentados na Figura 25.

Figura 25 - Estrutura do capítulo 6



Fonte: Adaptado de Pauwells *et al.* (2016).

6 APRESENTAÇÃO E ANÁLISES DE DADOS

Este capítulo apresenta os dados e os resultados das análises do reuso de dados secundários (*datasets*) (OE3) e da análise dos Recursos de Aprendizagem *Online* (OE4). Os dados secundários (*datasets*), relembra-se, foram localizados em fontes de dados abertos e os Recursos de Aprendizagem *Online* foram identificados a partir da indicação dos quadros conceituais recuperados na Revisão de Literatura, considerando que alguns são resultados das necessidades identificadas nos estudos que moldaram os *frameworks*, conforme já mencionado.

A partir dos requisitos identificados nos *frameworks*, foi possível consolidar as perspectivas e as dificuldades dos pesquisadores na prática do reuso dados de pesquisa sob os aspectos destes requisitos. A análise dos Recursos de Aprendizagem *Online* permitiu verificar a aderência dos requisitos e das competências decorrentes identificadas, sendo a base para a proposta conceitual de requisitos e competências essenciais para o reuso (OE5). Os requisitos propostos foram validados pelos critérios de melhores práticas da W3C (OE6), consórcio internacional de desenvolvimento de padrões para a World Wide Web.

6.1 RESULTADOS DAS ANÁLISES DOS CONJUNTOS DE DADOS SOBRE REÚSO DE DADOS.

Três conjuntos de dados (*datasets*) foram reutilizados neste estudo: Kim (2016 [*datasets*]), Kim (2017 [*datasets*]) e Berghmans *et al.* (2017 [*datasets*]). Após as análises separadas desses *datasets*, procedeu-se com a consolidação dos resultados, buscando aderência aos requisitos propostos. Em seguida, apresentam-se os resultados da análise dos dados secundários que foram reutilizados.

6.1.1 Resultados das Análises dos *Datasets* – conjunto de dados de Kim (2016 [*datasets*])

A pesquisa de Kim (2013), que originou o conjunto de dados de Kim (2016 [*datasets*]), foi utilizada neste estudo dentro da perspectiva de identificar as percepções de pesquisadores no compartilhamento de seus dados de pesquisa, permitindo a identificação de aspectos positivos e negativos que influenciam estruturas individuais, institucionais e, em sentido mais amplo, áreas de conhecimento. Tais percepções permitem inferir sobre aspectos normativos, estruturais e comportamentais para um conjunto de requisitos em dados.

O primeiro *dataset*, de Kim (2016 [*datasets*]), foi coletado para a investigação do comportamento de pesquisadores quanto ao compartilhamento de dados em suas áreas de atuação. Kim (2013) levantou os fatores institucionais e individuais que influenciavam o compartilhamento de dados por cientistas em diferentes áreas do conhecimento, como por exemplo: no nível individual Kim (2013; 2016 [*datasets*]) tratou da percepção dos benefícios auferidos à carreira, dos riscos relacionados, dos esforços despendidos no reuso de dados, do altruísmo no meio acadêmico e dos comportamentos de compartilhamento de dados, mas também das pressões exercidas pelas agências de fomento, pelos setores de publicação, pelos órgãos de regulação e mesmo pelos repositórios.

Em relação ao aspecto dos benefícios percebidos, os maiores índices positivos, identificados por Kim (2013, 2016), estão relacionados ao ganho de *status* para o pesquisador que compartilhar dados em sua própria comunidade acadêmica, ficando em torno de 51% das Ciências Sociais e 63% na demais Ciências pesquisadas. Nas Ciências Sociais, benefícios como recebimento de crédito acadêmico por meio de citação de dados compartilhados e benefícios extensíveis à carreira acadêmica, têm uma percepção positiva bem reduzida, em torno de 10% e 15% diante da positividade percebida nas demais disciplinas (KIM, 2016 [*datasets*]).

Percebe-se, portanto, a partir desses percentuais, estabelecidos mediante a análise das perguntas selecionadas para o escopo deste estudo (IDs), que esses pesquisadores não conhecem o aumento do impacto nas métricas de citação a partir do compartilhamento de dados de um pesquisador, por isso essa visão seria mais negativa a este respeito (IDs: *Benefit1*, *Benefit2*, *Benefit3* e *Benefit4*), fato também evidenciado por Curty (2016b), Piwowar e Vision (2013) e Sayão e Sales, (2016).

Em relação aos riscos percebidos, o maior receio dos pesquisadores, tanto das Ciências Sociais como das demais áreas, é de que seus dados possam ser utilizados incorretamente ou serem mal interpretados por outros pesquisadores. Sobre esse receio, deduz-se que um melhor conhecimento sobre a elaboração e a localização da documentação adjunta aos conjuntos de dados auxilie no entendimento do contexto e do teor das informações dos dados brutos, sendo assim possível minimizar este receio (IDs: *Risk1*, *Risk2*, *Risk3* e *Risk4*).

Índices como 61% para as Ciências Sociais e 65% para as demais áreas revelam que são necessários esforços significativos para o compartilhamento de dados. Do ponto de vista do reuso, essa percepção pode dificultar a disponibilidade de dados de maneira adequada e que viabilize o reuso (IDs: *Effort1*, *Effort2*, *Effort3* e *Effort4*).

Entretanto, índices em torno de 90% e 80% sinalizam com positividade o compartilhamento de dados, como sendo algo valioso e desejável, tanto nas Ciências Sociais

quanto nos demais Ciências. Portanto, infere-se que a falta de conhecimento sobre o tratamento dos dados ao longo do ciclo da pesquisa possa aumentar os esforços em preparar os dados para o compartilhamento ao final da mesma (IDs: *Attitude1*, *Attitude2*, *Attitude3* e *Attitude4*).

Percebe-se um índice aumentado de favorabilidade sobre o compartilhamento de dados dentro das áreas respectivas, tendo resultados muito próximos tanto nas Ciências Sociais (84%) como nas demais áreas (88%). De maneira geral, os pesquisadores percebem positivamente o compartilhamento de dados como um fator de contribuição para uma melhor pesquisa científica e de maior qualidade, fato este corroborado por Curty (2016a, 2016b); Curty *et al.* (2017); Sayão e Sales, (2014); Tenopir (2011) e Tenopir *et al.* (2015a). Em relação ao reúso, essa percepção pode contribuir com uma maior aceitação desta prática pela comunidade acadêmica disciplinar, como destacado por Curty, (2016a) e Tenopir *et al.* (2015a). (IDs: *AltruismCommunity1*, *AltruismCommunity2*, *AltruismCommunity3* e *AltruismCommunity4*).

Sob o aspecto de metadados⁴⁶, apenas 25% dos pesquisadores das Ciências Sociais consideram que o acesso às informações que identificam os conjuntos de dados seja fácil, enquanto que nas demais disciplinas⁴⁷, o percentual fica acima de 30%, chegando a 37% nas disciplinas citadas, segundo a percepção positiva sobre a disponibilidade de metadados para o compartilhamento. Ainda que este resultado esteja tratando do compartilhamento, os metadados também são úteis para o reúso, visto que é por meio deste conjunto de campos (metadados) que se pode conhecer o teor dos dados que estão sendo reutilizados, ou seja, os dados com os quais

⁴⁶ Metadados é um termo frequentemente usado para se referir a um tipo específico de documentação que possui uma estrutura definida e geralmente é projetada para ser legível por máquina - que descrevem aspectos significativos, por exemplo, conteúdo, contexto e estrutura das informações de um registro. (01-Bodleian\\Managing research data and data management planning\\Research-data-management-planning-introduction-handbook) (08-Managing and Sharing Research Data handbook\\6 STORING AND TRANSFERRING DATA\\CHAPTER 6 cd062-preservationpolicy).

⁴⁷ Aerospace Engineering, Agricultural Economics, Agricultural Engineering, Agricultural Sciences, Anatomy, Anesthesiology, Animal Sciences, Anthropology, Astronomy, Atmospheric Sciences, Biochemistry, Biology, Biomedical Engineering, Biometry and Epidemiology, Biophysics, Biosciences, Botany, Cardiology, Cell Biology, Chemical Engineering, Chemistry, Civil Engineering, Clinical Medicine, Clinical Psychology, Communication Disorders Sciences, Computer Engineering, Computer Science, Dental Sciences, Developmental Biology, Earth, Earth Atmospheric and Ocean Sciences, Ecology, Economics, Electrical Engineering, Endocrinology, Engineering, Engineering Science, Engineering Physics, Entomology and Parasitology, Environmental Engineering, Environmental Studies, Fishing and Fisheries Sciences, Food Sciences, Forestry, Gastroenterology, Genetics, Geography, Geosciences, Health Related, Hematology, History and Philosophy of Science, Industrial/Manufacturing Engineering, Linguistics, Mathematics and Applied Mathematics, Mechanical Engineering, Metallurgical and Materials Engineering, Microbiology Immunology and Virology, Microbiology Immunology and Virology, Molecular Biology, Natural Resources Conservation, Neurology, Neuroscience, Nuclear Engineering, Nursing, Nutrition, Obstetrics and Gynecology, Ocean Sciences, Oncology/Cancer Research, Ophthalmology, Pathology, Pediatrics, Pharmaceutical Sciences, Pharmacology, Physical Sciences, Physics, Physiology, Plant Sciences, Political Science, Preventive Medicine and Community Health, Psychiatry, Psychology, Public Administration, Pulmonary Disease, Radiology/Social Sciences, Sociology, Soil Sciences, Statistics, Surgery, Veterinary Sciences, Wildlife and Wildlands Sciences, Zoology.

se quer trabalhar. Sendo assim, infere-se que haja uma dificuldade relacionada com a identificação dos padrões de metadados por parte dos pesquisadores (IDs: *Metadata1*, *Metadata2* e *Metadata3*). As perguntas relacionadas aos metadados estão descritas na Tabela 14.

Tabela 14 - Perguntas relacionadas ao tema metadados

Seção	Pergunta	ID Pergunta	Ciências Sociais		Demais Disciplinas	
			N	%	N	%
			122		2.202	
Metadata	Researchers can easily access metadata.	Metadata1	30	25%	737	33%
	Metadata is available for researchers to share data.	Metadata2	26	21%	817	37%
	Researchers have the metadata necessary to share data.	Metadata3	21	17%	680	31%

Fonte: Apêndice L, adaptado de Kim (2016).

Há um consenso positivo, em torno de 60%, em relação a ideia de que os repositórios de dados sejam facilmente acessíveis e que estejam disponíveis para que os dados sejam compartilhados, tanto nas Ciências Sociais quanto nas demais disciplinas. Dessa forma, infere-se que os pesquisadores, cientes da disponibilidade de repositórios para o depósito de dados, também possam localizar nos mesmos repositórios, conjuntos de dados para serem reutilizados (IDs: *Repository1*, *Repository2* e *Repository3*).

Há, igualmente, um expressivo altruísmo em relação ao compartilhamento de dados, tanto do ponto de vista de auxílio aos outros pesquisadores com a disponibilidade de seus dados, chegando a uma resposta positiva de até 84% nas Ciências Sociais e 90% nas demais disciplinas, quanto na percepção de que a disponibilidade dos dados em si ajuda os demais pesquisadores em suas investigações. Dessa forma, infere-se que também haja uma percepção positiva de que os dados de outros pesquisadores possam auxiliar no desenvolvimento das próprias pesquisas dos entrevistados (IDs: *Altruism1*, *Altruism2* e *Altruism3*).

Sob o aspecto de frequência do compartilhamento, espaços públicos na Web ainda são mais utilizados do que os repositórios de dados disciplinares. A disponibilidade dos dados como material suplementar ainda tem uma grande representatividade como forma de provimento de acesso aos dados com 24% nas demais Ciências, sendo mais restrito nas Ciências Sociais, tendo apenas 4% apontado para esta opção. Essa pergunta acabou revelando um conflito, visto que no bloco anterior sobre a percepção de facilidade de acesso aos repositórios de dados, as respostas eram extremamente positivas, com índices acima de 60%.

Dessa forma, percebe-se que, ainda que os pesquisadores saibam onde acessar os repositório de dados, este não é o meio pelo qual eles mais compartilham os seus dados de pesquisa. A forma que os pesquisadores mais compartilham os seus dados ainda é via a comunicação pessoal como resposta aos requerentes, sendo esta a forma indicada por 12% nas Ciências Sociais e 26% nas demais Ciências (IDs: *MethodDisciplinaryRepository*, *MethodInstitutionalRepository*, *MethodPublicWeb*, *MethodSupplement*, *MethodRequested*, *MethodResponded*).

Em relação aos recursos disponíveis para promoção do compartilhamento dos dados de pesquisas, a percepção dos pesquisadores é bastante negativa. Apenas 14% das Ciências Sociais e 21% nas demais Ciências afirmou ter recebido uma instrução especializada. Apenas 12% nas Ciências Sociais afirmam ter uma pessoa especializada para auxiliar o seu compartilhamento de dados, enquanto que para as demais Ciências, este índice ficou em 17%. Há um ligeiro aumento da percepção positiva sobre tecnologias de informação disponíveis para auxiliar o compartilhamento de dados, em torno de 32% para as Ciências Sociais a 35% para as demais.

Enquanto que a percepção sobre o papel da organização dos pesquisadores em oferecer suporte ao compartilhamento de dados é mais negativa, pois apenas 16% nas Ciências Sociais e 25% nas demais disciplinas tem uma percepção positiva do suporte vindo de suas organizações. Nota-se que os pesquisadores também têm uma percepção mais negativa em relação aos recursos disponíveis para auxiliá-los (IDs: *Resource1*, *Resource2*, *Resource3* e *Resource4*).

Essas percepções se dão sob a ótica do compartilhamento de dados, que frequentemente é uma prática mandatória por força de políticas de governos às quais os pesquisadores estão submetidos por receber financiamento (EUROPEAN COMMISSION, 2016a; KILEY, 2013; NATIONAL SCIENCE FOUNDATION, 2007). Portanto, se para o compartilhamento de dados não há uma instrução especializada, infere-se que para o reuso haja uma percepção ainda mais negativa dos pesquisadores, visto que não há qualquer tipo de obrigatoriedade nesta prática.

A partir dessas análises, destacam-se as necessidades de capacitação e apoio aos pesquisadores visando o reuso de dados, a respeito dos sentimentos positivos e negativos identificados:

a) Sentimentos positivos identificados:

- repositório de dados estão facilmente acessíveis;
- a disponibilidade dos dados auxilia os demais pesquisadores em suas pesquisas;

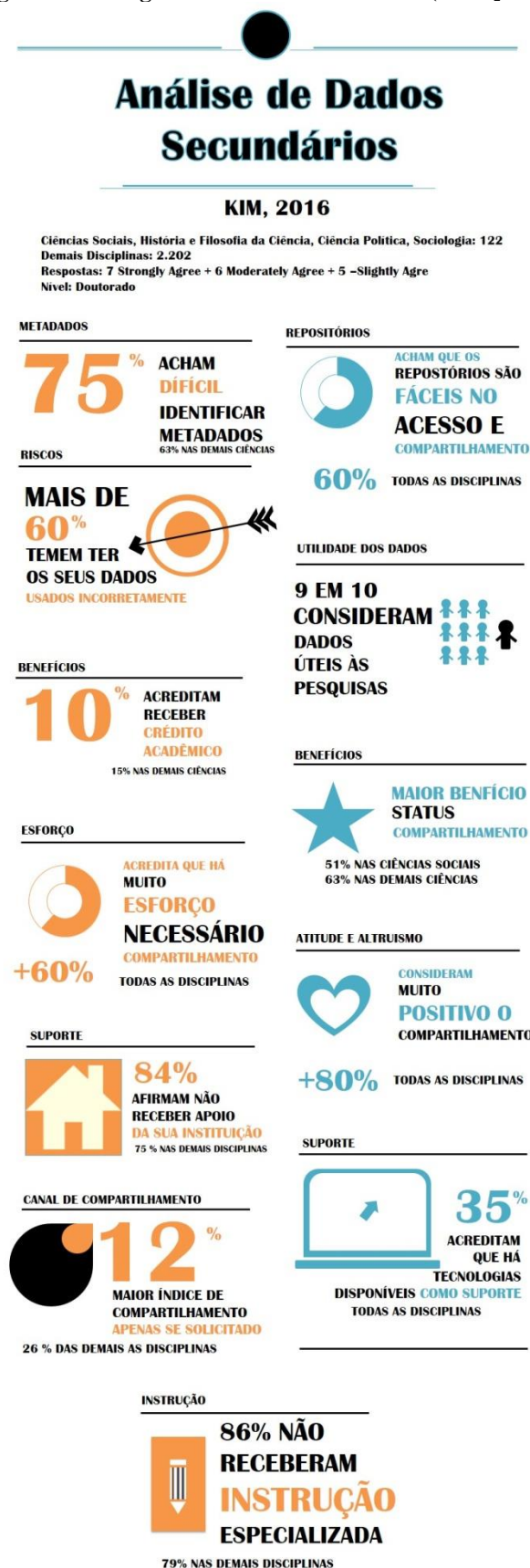
- em relação ao reúso, a percepção positiva sobre o auxílio dos dados secundários em pesquisas pode contribuir com uma maior aceitação desta prática pela comunidade acadêmica disciplinar;
- tecnologias de informação estão disponíveis para o compartilhamento de dados, sendo os recursos que mais auxiliam os pesquisadores.

b) Sentimentos negativos: oportunidades de letramento

- identificação dos padrões de metadados;
- impacto nas métricas de citação a partir do compartilhamento de dados de um pesquisador;
- elaboração e a localização da documentação adjunta aos conjuntos de dados minimizando risco de mau uso;
- conhecimento sobre o tratamento dos dados ao longo do ciclo da pesquisa;
- vantagens do uso dos repositórios de dados disciplinares como meio de compartilhamento de dados;
- identificar e localizar os recursos disponíveis para auxílio, como treinamentos online, por exemplo.

O resumo das análises dos pontos positivos e negativos do conjunto de dados de Kim, (2016 [*datasets*]) está representado no Infográfico da Figura 26.

Figura 26 - Infográfico das análises de Kim (2016 [datasets])



Fonte: Autoria própria (2019)

Legenda:

Negativo

Positivo

As principais características do conjunto de dados de Kim (2016 [*datasets*]), além das perguntas selecionadas e dos resultados extraídos para a análise das respostas, estão disponíveis no Apêndice L, na Tabela 21.

6.1.2 Resultados das Análises dos *Datasets* – conjunto de dados Kim (2017 [*datasets*])

Kim (2017 [*datasets*]), realizou outro estudo com o foco no comportamento do pesquisador quanto ao reuso de dados de pesquisa em várias disciplinas⁴⁸ do conhecimento. A investigação de Kim (2017 [*datasets*]), em relação ao reuso de dados, traz no primeiro bloco de perguntas selecionadas a percepção da utilidade do reuso de dados secundários. De maneira geral, todas as perguntas em relação à utilidade demonstram uma percepção positiva, acima de 50%. A questão de maior relevância positiva está relacionada com a melhoria percebida da eficácia da própria pesquisa a partir do reuso de dados secundários, sendo um índice positivo de 57% nas Ciências Sociais e de 61% nas demais Ciências. Desta forma, registra-se que os pesquisadores percebem utilidade na prática do reuso de dados (IDs: *Usefulness1*, *Usefulness2* e *Usefulness3*).

Entretanto, no próximo bloco de perguntas, relacionadas com as preocupações, identifica-se que o maior receio ainda está na má interpretação dos dados, sendo o índice de concordância com esse risco em torno de 38% para todas as Ciências. A preocupação em causar algum tipo de infração ao se reusar os dados cai para 17% nas Ciências Sociais e para 23% das demais Ciências investigadas neste estudo.

Se comparado ao estudo anterior (KIM, 2016), sobre o compartilhamento de dados, o quesito de risco de má interpretação dos dados alcançou 61% nas Ciências Sociais e 63% nas demais Ciências. Sendo assim, os pesquisadores parecem ter maior receio de que seus dados sejam utilizados de forma equivocada, demonstrando mais confiança de que eles próprios utilizariam os dados secundários de outrem de forma mais assertiva. Nessa questão, infere-se que é necessário reforçar o entendimento da documentação de apoio aos conjuntos de dados, tanto do ponto de vista do compartilhamento, em que os pesquisadores precisarão preparar a documentação referente aos seus *datasets*, como também do ponto de vista do reuso, pois eles necessitarão compreender o conteúdo e as instruções desta documentação no momento em que se pretender efetivar o reuso (IDs: *Concern1*, *Concern2* e *Concern3*).

⁴⁸ Agricultural Sciences, Biological Sciences, Computer Sciences, Earth, Atmospheric, and Ocean Sciences, Engineering, Health Sciences, Physical Sciences, Psychology e Social Sciences.

Em relação aos esforços na localização e no acesso aos conjuntos de dados, além do processamento dos dados para um novo estudo, estas são vistas como atividades de expressiva alocação de energia. Mais de 50%, variando de 52% a 57% dos pesquisadores nas Ciências Sociais percebem essas atividades como uma grande demanda de esforços. Nas demais Ciências, esse índice cresce, ficando o nível de percepção de esforços a serem empregados na localização, acesso e reuso dos dados disponíveis em torno de 65% a 68% (IDs: *Effort1*, *Effort2* e *Effort3*).

Diante da questão que aborda os repositórios de dados, ainda que em torno de 60% dos pesquisadores afirmem que os repositórios estejam disponíveis, um percentual menor, em torno de 47% deles, afirmam que esses repositórios podem ser facilmente acessados para reusar dados. Nessa questão, infere-se que os pesquisadores tenham acesso aos repositórios, porém, podem enfrentar alguma dificuldade em localizar dados válidos ou de qualidade para o reuso, ou ainda dificuldades nas atividades de coleta e curadoria desses dados para viabilidade do reuso (IDs: *Repository1* e *Repository2*).

Em termos de recursos para apoio à atividade de reuso de dados, há uma percepção de que tecnologias de informação estão disponíveis para suporte aos pesquisadores, sendo este índice positivo em 51% nas Ciências Sociais e 45% nas demais Ciências (IDs: *OrgResource1*, *OrgResource2* e *OrgResource3*). No estudo anterior de (Kim (2016 [*datasets*])), uma pergunta que também abrangia a disponibilidade de recursos para auxílio aos pesquisadores no compartilhamento de dados, também teve resposta com uma percepção positiva, em torno de 32% a 35%, de que haviam tecnologias de informação disponíveis para essa finalidade (KIM (2016 [*datasets*]), ID *Resource4*).

Corroborando este estudo à visão da falta de apoio de pessoas especializadas para auxílio nas questões de reuso de dados, sendo esta uma consciência com apenas 21% de positividade nas Ciências Sociais e 15% nas demais Ciências. Esses são índices muito próximos de uma pergunta do mesmo teor no estudo anterior (Kim (2016 [*datasets*])). Pessoas ou grupos especializados para auxiliar em relação ao compartilhamento e/ou ao reuso de dados são os recursos com a menor percepção positiva dentre os pesquisadores em ambas investigações. (KIM, 2016 [*datasets*]), ID *Resource4*); (KIM, 2017 [*datasets*]), ID *Resource3*).

Neste sentido, percebe-se uma lacuna para que as bibliotecas acadêmicas, que já ofertam serviços de apoio aos pesquisadores nas suas atividades científicas, possam também suprir essa ausência de pessoas preparados para auxiliar, capacitar os pesquisadores em gerenciar os seus próprios dados, além de também prestarem um novo serviço relevante à necessidade da pesquisa acadêmica do século XXI.

No bloco de perguntas sobre as intenções para o reúso, mais de 50% dos pesquisadores de todas as disciplinas gostariam de utilizar dados secundários para suas próprias pesquisas no futuro. Em torno de 42% deles registram que têm realmente a intenção e/ou que irão tentar executar essa prática. Ou seja, considerando que o reúso de dados não é uma atividade mandatária como é por vezes o compartilhamento de dados, revela-se uma quantidade expressiva de pesquisadores que está disposta a reusar dados em suas pesquisas no futuro. Isso reforça a relevância para que requisitos e critérios de letramento em dados específicos para o reúso sejam identificados e disponibilizados, nos quais os pesquisadores possam aprender e se beneficiar neste processo, minimizando os esforços requeridos e ampliando a vantagem percebida da atividade do reúso (IDs: *Intention1*, *Intention2* e *Intention3*). Em relação aos sentimentos positivos e negativos, tem-se:

a) Sentimentos positivos identificados:

- reúso de dados secundários melhora a eficácia da própria pesquisa;
- repositórios estão disponíveis, porém, podem não ser tão facilmente acessados para reusar dados;
- uma percepção de que tecnologias de informação estão disponíveis para suporte aos pesquisadores, tanto para compartilhamento quanto para reúso;
- quase metade dos pesquisadores realmente têm a intenção e/ou irão tentar executar o reúso de dados.

b) Sentimentos negativos: oportunidades de letramento

- maior receio ainda está na má interpretação dos dados, seguido do medo de causar algum tipo de infração ao se reusar os dados;
 - melhor entendimento da documentação de apoio que instrui os conjuntos de dados;
 - maior compreensão das regras de propriedade intelectual, direitos autorais dos dados e licenças de uso e reúso;
- percepção de que grandes esforços em geral são empregados na localização, acesso e reúso dos dados disponíveis;
 - capacitação nas atividades de coleta e curadoria desses dados para viabilizar o reúso;
 - como localizar dados válidos ou de qualidade para o reúso;
- falta de apoio de pessoas especializadas para auxílio nas questões de reúso de dados:
 - lacuna para que as bibliotecas acadêmicas possam suprir a ausência de pessoas preparados para auxiliar, capacitar os pesquisadores em gerenciar os seus próprios dados.

O resumo das análises dos pontos positivos e negativos do conjunto de dados de Kim, (2017 [datasets]), está registrado no Infográfico da Figura 27.

Figura 27 - Infográfico das análises de Kim (2017 [datasets])



Fonte: Autoria própria (2019)

Legenda:

Negativo

Positivo

Os resultados da análise do conjunto de dados de Kim, (2017 [*datasets*]), estão disponíveis íntegra no Apêndice M, na Tabela 22. Em seguida apresentam-se as análises do conjunto de dados de Berghmans *et al.* (2017 [*datasets*]).

6.1.3 Resultados das Análises dos *Datasets* – conjunto de dados Berghmans *et al.* (2017 [*datasets*])

Berghmans *et al.* (2017 [*datasets*]), avaliaram um amplo conjunto de dados de pesquisadores de diversas disciplinas⁴⁹ e em diferentes países sobre os hábitos e as percepções nas práticas diárias de pesquisa. Enquanto os estudos de Kim (2016; 2017 [*datasets*]), restringiram-se aos pesquisadores norte-americanos, o conjunto de dados de Berghmans *et al.* (2017 [*datasets*]) envolveu pesquisadores de diversos países, o que motivou a seleção destes dados pela perspectiva global das percepções. O estudo contrapõe a expectativa de que a força política atue sobre a ciência e os dados abertos, argumentando que a mudança prática está muito mais limitada no fato de que as recompensas são poucas e a complexidade em disponibilizar dados é alta. Os dados deste estudo, conduzido pela Leiden University e Elsevier (2017), revelam que entre as interfaces da política e da prática, são necessários mais esforços para tornar os dados abertos uma ação para a inovação na pesquisa (LEIDEN UNIVERSITY; ELSEVIER, 2017).

A primeira questão investigada demonstra que as Ciências Sociais fazem uso intensivo de dados empíricos e observacionais, como questionários, entrevistas e dados etnográficos, chegando a um índice de seleção de 87% deste tipo de dados, frente a 54% desta escolha pelas demais áreas (IDs: Q1a_1 a Q1a_5).

Em relação aos formatos de dados utilizados, os tipos numéricos ainda são mais recorrentes, sendo estes apontados como escolha de 80% dos pesquisadores nas Ciências Sociais e 67% nas demais Ciências. A segunda opção mais utilizada em todas as Ciências foi o formato de arquivo textual (IDs: Q1b_1 a Q1b_8).

Em relação aos tipos de dados a serem compartilhados, os pesquisadores também apontaram os formatos para tratamentos quantitativos com o uso de aplicativos como Excel e sistemas de análise de dados estatísticos como sua principal escolha. Portanto, depreende-se

⁴⁹ Agriculture, Arts and Humanities, Astronomy, Biochemistry, Genetics and Molecular Biology, Biological Sciences, Chemical Engineering, Chemistry, Computer Sciences / IT, Earth and Planetary Sciences, Economics, Electrical / Electronic Engineering, Engineering and Technology, Environmental Sciences, Immunology and Microbiology, Materials Science, Mathematics, Medicine and Allied Health, Neuroscience, Nursing, Pharmacology, Toxicology and Pharmaceuticals, Physics, Psychology e Social Science.

que habilidades em lidar com planilhas e sistemas de análise de dados são relevantes para todos os pesquisadores (IDs: Q1c_1 a Q1c_8).

Com respeito à propriedade intelectual sobre os dados, um índice de 63% nas Ciências Sociais e de 61% nas demais Ciências apontou os próprios pesquisadores como proprietários dos dados, antes e também depois de sua publicação. Em segundo lugar como proprietários dos dados antes da publicação, estão indicados os colegas e os colaboradores das pesquisas (IDs: Q1ei_1 a Q1ei_99). Ainda sobre a investigação da propriedade dos dados após a publicação, os pesquisadores mantiveram a indicação de si próprios como os que detém a propriedade intelectual dos dados de pesquisa, sendo essa a escolha de 48% dos pesquisadores nas Ciências Sociais e 44% para as demais disciplinas. Os pesquisadores das Ciências Sociais também atribuem a propriedade após a publicação dos dados para os financiadores de pesquisa, com o índice de 35%, enquanto 39% dos pesquisadores das demais disciplinas indicaram suas instituições como proprietárias dos dados (IDs: Q1eii_1 a Q1eii_99). Nesse aspecto, aprofundamento nos conceitos sobre Ciência Aberta e os demais movimentos que a impulsionam (como acesso e dados abertos) são relevantes para uma melhor compreensão de *copyright* e licenças (CARLSON *et al.*, 2011; CARLSON; JOHNSTON, 2015; FACILITATE OPEN SCIENCE TRAINING FOR EUROPEAN RESEARCH, 2015; SHERPA RoMEO, *online*; TASK GROUP ON DATA CITATION STANDARDS AND PRACTICES, 2013).

Como forma de publicação de dados, o repositório de dados ainda é visto como principal opção nas Ciências Sociais, com 19% de escolha. Já nas outras disciplinas, os dados devem ser publicados em forma de artigo de dados separado, com avaliação por pares 29% (IDs: Q1f_1 a Q1f_3).

Para 67% dos pesquisadores nas Ciências Sociais e 73% dos pesquisadores nas demais disciplinas, acessar os dados de outros pode beneficiar sua própria pesquisa. Entretanto, apenas 27% nas Ciências Sociais e 24% nas demais disciplinas afirmam terem recebido treinamento suficiente para compartilhar os seus dados. Um percentual acima de 55% afirma que é necessário que especialistas em gestão de dados exerçam seu papel neste cenário (IDs: Q2a_1 a Q2a_10).

Nos benefícios percebidos sobre a publicação dos dados em conjunto com o artigo, os cientistas sociais indicaram a reprodução da pesquisa, com 19% de indicações, como sendo o maior benefício esperado. Em segundo lugar ficou a maior possibilidade de colaboração, com 16%, acima ainda da possibilidade de reuso de dados que obteve 15% de indicação. Já nas demais disciplinas, a possibilidade de colaboração foi a primeira opção com de 30% de escolha,

a reprodução da pesquisa foi indicada logo atrás com 28%. O reuso de dados também foi selecionado como o terceiro benefício, com 20% das indicações (IDs: Q2c_1 a Q2c_99).

Na questão gestão de dados, 49% dos cientistas sociais e 45% das demais disciplinas tomam algum tipo de medida na gestão ou no arquivamento dos seus dados para viabilizar o reuso. Em relação aos que nunca fazem esse tipo de ação, um índice de 16% foi registrado nas ciências sociais, enquanto que nas demais ciências apenas 1% afirmou não preparar seus dados (IDs: Q2e_1 a Q2e_3).

Em relação ao uso de atribuição de licenças no compartilhamento de dados, 60% de todos os respondentes não sabem qual o tipo de licença escolher, enquanto que a segunda opção apontou a licença com as maiores restrições dentro da classificação de Creative commons (IDs: Q3d_1 a Q3d_99). Isso demonstra que conceitos de abertura de dados, *copyright* e *copyleft* e como mensurá-los podem auxiliar os pesquisadores tanto a atribuir permissões aos seus dados quanto a identificar de que forma os dados secundários podem ser reutilizados.

Com relação a confiança nos dados secundários de outros pesquisadores, os fatores mais importantes indicados por aproximadamente 40% dos pesquisadores estão na documentação completa e nos metadados adequados, além do cumprimento de requisitos sobre as orientações de formatos dos arquivos. Nota-se que questões de documentação adequada são mais relevantes enquanto critério de confiabilidade do que questões relacionadas à reputação de uma instituição, de um repositório, ou de uma pessoa renomada (IDs: Q3e_1 a Q3e_9).

Sobre as regras de citação, 57% dos pesquisadores nas Ciências Sociais e 46% dos pesquisadores das demais disciplinas indicaram que não há padrões para a citação de dados de pesquisa nas suas respectivas disciplinas. Isso alerta para a necessidade de maior familiaridade dos pesquisadores com os padrões de citação de dados já desenvolvidos e consolidados por entidades relacionadas à gestão de dados científicos (IDs: Q3f_1 a Q3f_4).

Compilando os sentimentos positivos e negativos tem-se:

a) Sentimentos positivos identificados:

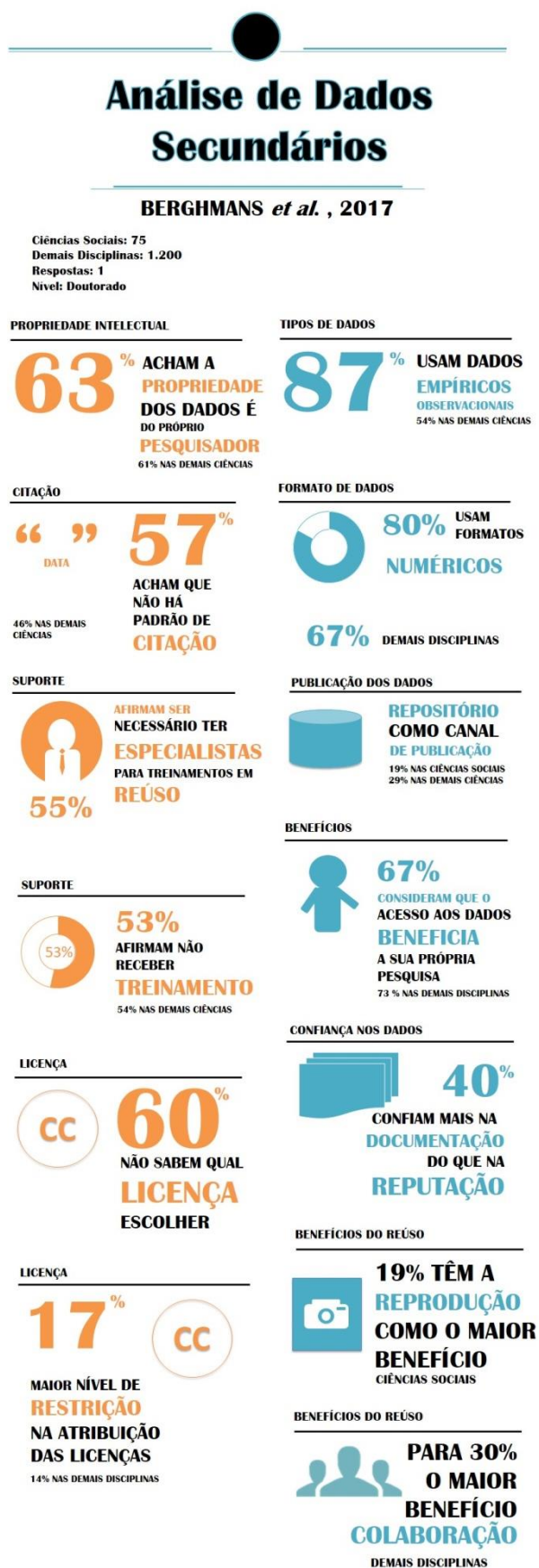
- Ciências Sociais fazem uso intensivo de dados empíricos e observacionais, predominantemente qualitativos;
- formatos de tipos numéricos são os mais recorrentes, tanto para uso quanto para compartilhamento;
- o repositório de dados é considerado o meio de compartilhamento dos dados;
- acessar os dados de outros pode beneficiar sua própria pesquisa;

- reprodução da pesquisa é o maior benefício esperado no compartilhamento, seguido da possibilidade de colaboração e do reuso de dados;
- metade dos pesquisadores fazem alguma medida para viabilizar o reuso dos seus dados;
- 40% da confiança dos pesquisadores nos dados está na documentação completa e nos metadados adequados.

b) Sentimentos negativos: oportunidades de letramento

- maior demanda habilidades em lidar com planilhas e sistemas de análise de dados numéricos para todos os pesquisadores;
- ampla maioria apontou os próprios pesquisadores como proprietários dos dados, antes e também depois de sua publicação;
 - Familiaridade com conceitos de Ciência, acesso e dados abertos;
- maior demanda habilidades em lidar com planilhas e sistemas de análise de dados;
- apenas um terço afirma ter recebido treinamento suficiente para compartilhar os seus dados:
 - mais da metade afirma que é necessário que especialistas em gestão de dados exerçam seu papel neste cenário;
 - incorporar atividades de capacitação como rotina aos pesquisadores;
- 60% de todos os respondentes não sabem qual o tipo de licença escolher:
 - conhecimento da classificação de licença Creative Commons;
- maior percepção dos pesquisadores de que não há padrões para a citação de dados de pesquisa nas suas respectivas disciplinas:
 - acesso às opções de padrões de citação estabelecidos na comunidade internacional.

O resumo das análises dos pontos positivos e negativos do conjunto de dados de Berghmans *et al.* (2017 [*datasets*]) é demonstrado na Figura 28.

Figura 28 - Infográfico das análises de Berghmans *et al.* (2017 [datasets])

Fonte: Autoria própria (2019)

Legenda:

Negativo

Positivo

Os resultados completos da análise do conjunto de dados de Berghmans *et al.* (2017 [datasets]), estão disponíveis no Apêndice N - na Tabela 23.

A seguir, após as análises dos dados secundários separadamente, as principais dificuldades identificadas no reúso dados de pesquisa são consolidadas e agrupadas, sendo relacionadas com cada um dos requisitos de reúso.

6.1.4 Consolidação dos Resultados das Análises dos *Datasets*

Esta seção visa consolidar as perspectivas e as dificuldades dos pesquisadores identificadas no reúso de dados de pesquisa, por meio de dados secundários (*datasets*), conforme previsto no Objetivo Específico 3 (OE3). A análise dos dados secundários evidenciou pontos positivos sobre as perspectivas dos pesquisadores, assim como oportunidades de letramento para a melhoria do reúso de dados, a partir de dificuldades que os pesquisadores apontaram nos questionários avaliados. Esses aspectos estão previamente detalhados nas seções 6.1.1, 6.1.2 e 6.1.3.

Porém, com intuito de consolidar essas informações, cada aspecto foi relacionado com os requisitos aos quais eram mais aderentes dentre os identificados na análise dos *frameworks*. Dessa forma, é possível identificar quais detalhamentos são necessários e as competências mais relevantes para a realidade dos pesquisadores nas Ciências Sociais.

Foi percebida uma lacuna para acomodar um conjunto de demandas dos pesquisadores, portanto, criou-se um novo requisito não presente nos *frameworks*, denominado “*Support and formal training*” (suporte e treinamento formal). As perspectivas e as dificuldades dos pesquisadores no reúso de dados estão relacionadas no Quadro 21. Bem como, quais competências foram possíveis identificar nos dados secundários e que podem ser atribuídas a esses requisitos e como o letramento pode favorecer ou minimizar essas dificuldades e/ou percepções dos pesquisadores. Após o quadro supramencionado apresenta-se a análise dos recursos de aprendizagem *online*, que identifica as principais competências para o reúso de dados disponíveis nos ambientes virtuais de aprendizagem selecionados; conforme previsto no Objetivo Específico 4 (OE4).

Quadro 21 - OE3: Perspectivas e dificuldades dos pesquisadores no reúso de dados – análise dos dados secundários

(continua)

Categorias de Análise	Unidades de Registro	Percepções e dificuldades dos Pesquisadores e competências necessárias			Contribuição do letramento em dados
		Kim (2016, [datasets])	Kim (2017, [datasets])	Berghmans <i>et al.</i> , (2017 [datasets])	Competências
1. Access	1.1 Access data sources	- Repositório de dados estão facilmente acessíveis para depósito - Vantagens do uso dos repositórios de dados disciplinares	Repositórios estão facilmente disponíveis para reúso	Repositório reconhecido como meio de compartilhamento dos dados	Conhecimento e uso de indexadores de repositórios de dados
	1.2 Discovery	n/a	Grandes esforços na localização de dados Dificuldade em localizar dados válidos	n/a	Uso de estratégia de pesquisa e familiaridade com motor de busca indexadores de repositórios de dados
	1.3 Access data	n/a	Grandes esforços no acesso aos dados	n/a	Conhecimento dos critérios de acesso, controlado ou não, aos dados
2. Analysis	2.1 Reading	n/a	n/a	n/a	n/a
	2.2 Interpreting/ Understanding data	Maior receio ao reúso ainda está na má interpretação dos dados	Maior receio ao reúso ainda está na má interpretação dos dados	n/a	Entendimento dos aspectos metodológicos e escopo das pesquisas. Acesso e leitura da documentação de dados
	2.3 Evaluating	n/a	n/a	n/a	n/a
	2.4 Critical thinking	n/a	n/a	n/a	n/a
3. Acquisition	3.1 Obtaining data	n/a	n/a	n/a	n/a
	3.2 Collecting Data	n/a	Coleta para viabilizar o reúso	n/a	Salvar e manter o acesso aos dados armazenados, de forma organizada com a nomeação e estruturação dos arquivos

(continua)

Categorias de Análise	Unidades de Registro	Percepções e dificuldades dos Pesquisadores e competências necessárias			Contribuição do letramento em dados
		Kim (2016, [datasets])	Kim (2017, [datasets])	Berghmans <i>et al.</i> , (2017 [datasets])	Competências
4. Archiving	4.1 Plan for archiving	n/a	n/a	n/a	n/a
	4.2 Archiving practices	n/a	n/a	n/a	n/a
	4.3 Data Preservation Techniques	n/a	n/a	n/a	n/a
5. Curation	5.1 Conversion	n/a	Curadoria para viabilizar o reuso	Uso intensivo de dados empíricos e observacionais	Domínio das metodologias e ferramentas de análises de dados empíricos e observacionais
	5.2 Formats	n/a		Uso intensivo de formatos de tipos numéricos	Conhecimento de softwares que permitem a gestão de dados numéricos (Excel, OpenOffice, Stata, Banco de dados, SPSS, Python, R).
	5.3 Merging	n/a		Planilhas e sistemas de análise de dados numéricos	
	5.4 Processing	n/a		n/a	n/a
	5.5 Synthesize	n/a		n/a	n/a
6. Data Management	6.1 Data Management Pela	Tratamento dos dados ao longo do ciclo da pesquisa	n/a	n/a	Conhecimento dos processos de ciclo de gestão de dados ao longo do desenvolvimento das pesquisas
	6.2 Directory structures	n/a	n/a	n/a	n/a
	6.3 Naming conventions	n/a	n/a	n/a	n/a
	6.4 Organizational planning	n/a	n/a	n/a	n/a
	6.5 Standards	n/a	n/a	n/a	n/a

(continua)

Categorias de Análise	Unidades de Registro	Percepções e dificuldades dos Pesquisadores e competências necessárias			Contribuição do letramento em dados
		Kim (2016, [datasets])	Kim (2017, [datasets])	Berghmans <i>et al.</i> , (2017 [datasets])	Competências
7. Documentation	7.1 Data sets description	Localização da documentação	n/a	Documentação como requisito de confiança dos pesquisadores nos dados	Localizar e analisar a documentação disponível
	7.2 Documentation Methods	Elaboração da documentação	Melhor entendimento da documentação	Viabilizar o reuso dos seus dados	
8. Licences	8.1 Apply your licence	n/a	Maior compreensão das regras de licenças de uso e reuso	Creative Commons	Conhecimento dos padrões de licenças Creative Commons
	8.2 Citation	Impacto nas métricas de citação	n/a	Padrões de citação	Conhecimento dos indicadores de citação, de indicadores alternativos e das métricas de cálculos
	8.3 Intellectual Property Rights	n/a	Maior compreensão das regras de propriedade intelectual	Pesquisadores como proprietários dos dados	Lei de propriedade intelectual, Lei de direito autoral
	8.4 Licenses types and copyright	n/a	Maior compreensão das regras de direitos autorais e licenças dos dados	Creative Commons / pesquisadores como proprietários dos dados	Políticas de financiamento que regulam a propriedade dos dados,
	8.5 Publisher rights	n/a	n/a	Creative Commons / pesquisadores como proprietários dos dados	Políticas de editores que transferem a propriedade intelectual
	8.6 Using data ethically	n/a	n/a	n/a	Políticas de reuso
9. Metadata	9.1 Metadata Description	n/a	n/a	Confiança dos pesquisadores nos dados	Conhecimento dos conteúdos requeridos em metadados a fim de verificar a completude do preenchimento dos mesmos
	9.2 Metadata Standards	Identificação dos padrões de metadados	n/a	n/a	

(continua)

Categorias de Análise	Unidades de Registro	Percepções e dificuldades dos Pesquisadores e competências necessárias			Contribuição do letramento em dados
		Kim (2016, [datasets])	Kim (2017, [datasets])	Berghmans <i>et al.</i> , (2017 [datasets])	Competências
10. Open Science	10.1 Concept of Open Science	n/a	Melhora a eficácia da própria pesquisa	Conceitos de Ciência aberta	Familiaridade com os preceitos da Ciência Aberta
	10.2 Open Access	n/a	n/a	Conceitos de acesso aberto	Valor percebido do Acesso aberto
	10.3 Open Data Journals	n/a	n/a	Meio relevante de publicidade de dados	Conhecimento do papel e da disponibilidade dos periódicos de dados
	10.4 Open Data Standards	n/a	n/a	Conceitos de dados abertos	Reconhecer e utilizar padrões de viabilidade de dados abertos
	10.5 Open Data use and reuse	n/a	Grandes esforços no reúso Melhora a eficácia da própria pesquisa	Benefício esperado ao acessar dados	Conhecer estudos de sucesso em reúso de dados / identificação de fraudes pela reprodução
11. Policy	11.1 Funding policies	n/a		n/a	n/a
12. Quality	12.1 Quality Assurance	n/a	Localizar dados de qualidade para o reúso	n/a	Reconhecer padrões e selos de qualidade de repositórios, critérios de proveniência dos dados
	12.2 Research methods – practical application	n/a	n/a	n/a	n/a
	12.3 Research methods – theoretical knowledge	n/a	n/a	n/a	n/a

(conclusão)

Categorias de Análise	Unidades de Registro	Percepções e dificuldades dos Pesquisadores e competências necessárias			Contribuição do letramento em dados
		Kim (2016, [datasets])	Kim (2017, [datasets])	Berghmans <i>et al.</i> , (2017 [datasets])	Competências
13. Reuse	13.1 Reproducible Research	n/a	n/a	Benefício esperado ao acessar dados	Conhecer estudos de sucesso em reuso de dados / identificação de fraudes pela reprodução
	13.2 Legal requirements	n/a	Maior compreensão das regras de reuso	n/a	Políticas de reuso de dados
	13.3 Re-use policies	n/a	Quase metade dos pesquisadores realmente têm a intenção de reuso	n/a	
14. Sharing	14.1 Re-sharing	n/a	n/a	n/a	n/a
15. Storage	15.1 Backup	n/a	n/a	n/a	n/a
	15.2 Security	n/a	n/a	n/a	n/a
16. Using	16.1 Databases	n/a	n/a	n/a	n/a
	16.2 Data collections	n/a	Podem não ser tão facilmente acessados para reusar dados	n/a	Conhecer critérios de acesso, acesso controlado, regras de embargo de dados
17. Visualization	17.1 Visualization tools	n/a	n/a	n/a	n/a
18. Support and formal training	18.1 Support and formal training	- Identificar e localizar os recursos disponíveis para auxílio - Tecnologias de informação são recursos que mais auxiliam os pesquisadores	- Apoio de pessoas especializadas - oportunidade para as bibliotecas acadêmicas - Tecnologias de informação estão disponíveis para suporte	Treinamentos com especialistas em gestão de dados	Conhecer os programas de capacitação e treinamento das Universidades e instituições sobre Gestão de dados de Pesquisas

Fonte: Autoria própria (2019) a partir do reuso de dados dos *datasets* de Kim (2016 [datasets]), Kim (2017 [datasets]), Berghmans *et al.*, (2017 [datasets]).

6.2 RESULTADOS DAS ANÁLISES DOS RECURSOS DE APRENDIZAGEM *ONLINE*

Esta seção apresenta inicialmente os resultados coletados para as posteriores análises dos Recursos de Aprendizagem; em seguida são apresentados os resultados dos requisitos para reuso. As análises atendem ao propósito de identificar se os requisitos das categorias de análises e as competências das unidades de registro, levantadas previamente nos *frameworks*, estavam presentes nos recursos de aprendizagem *online*. Esta seção apresenta os resultados das análises primeiramente em relação a todos os requisitos (para reuso ou não) e, em seguida, os resultados das análises para os requisitos segmentados do reuso de dados.

6.2.1 Apresentação dos Resultados dos Recursos de Aprendizagem - Resultados Gerais

Inicialmente fez-se a codificação dos recursos de aprendizagem *online* com o *Software NVivo*. Selecionaram-se os arquivos a serem codificados e então o comando de Auto codificar foi executado. Um exemplo do recurso 01 - *data, management e research* é apresentado na Figura 29; salienta-se que foram gerados códigos automáticos para cada recurso, separadamente. Os resultados de codificação automática no *Software NVivo*, gerados a partir da codificação automática, no entanto, foram bem reduzidos. A codificação automática apresentou problemas de nomenclatura dos termos devido à algumas características dos arquivos analisados, conforme detalhado sequencialmente.

Figura 29 - Recursos de Aprendizagem *Online* – codificação das categorias de análise

The screenshot shows the NVivo software interface. The top menu bar includes 'Arquivo', 'Início', 'Importar', 'Criar', 'Explorar', and 'Compartilhar'. Below the menu is a toolbar with various icons for different functions like 'Projeto', 'NCapture', 'Arquivos', 'Excel', 'Arquivo de texto (.txt, .csv)', 'SurveyMonkey', 'Qualtrics', 'SPSS', 'Planilha de classificação', 'EndNote', 'Mendeley', 'RefWorks', and 'Zotero'. The main workspace is divided into a left sidebar and a central table. The sidebar shows a tree view of 'Recursos Aprendizagem' with '01-Bodleian' selected. The central table displays the automatic coding results for '01-Bodleian'.

Nome	Arquivos	Referências
data		14
management		12
research		12

A red arrow points from the '01-Bodleian' folder in the sidebar to the table. A blue callout box with the text 'Nós classificados automaticamente' points to the 'data', 'management', and 'research' rows in the table.

Fonte: Autoria própria com codificação automática com o *software NVivo* (2019).

Os arquivos desta pesquisa foram utilizados de acordo com a disponibilidade dos recursos de aprendizagem nos sítios utilizados. Entretanto, o *Software NVivo* – versão 12 Plus, não suporta arquivos em formato de apresentação PowerPoint, estes tiveram que ser convertidos em arquivos com extensão pdf. Essa conversão pode ter resultado em arquivos suscetíveis às falhas de leitura de palavras nos documentos, como revela a Figura 30, gerando, por exemplo, a palavra ‘ion’ como se fosse um termo novo, porém, esta era a extensão das palavras ‘preservation’ e ‘organization’ que o sistema registrou de forma separada.

Figura 30 - Exemplos de problemas com codificação automática

The screenshot displays the NVivo software interface with several annotations highlighting automatic coding issues:

- Problemas de nós automáticos**: A callout box pointing to the 'ion' node in the table, which is a result of the software incorrectly splitting the words 'organization' and 'preservation'.
- Nós automáticos**: A callout box pointing to the 'ion' node in the project tree on the left.
- Problemas com Nós automáticos**: A callout box pointing to the '6. Preservation Techniques' module in the document viewer, which is a result of the software incorrectly splitting the words 'preservation' and 'techniques'.

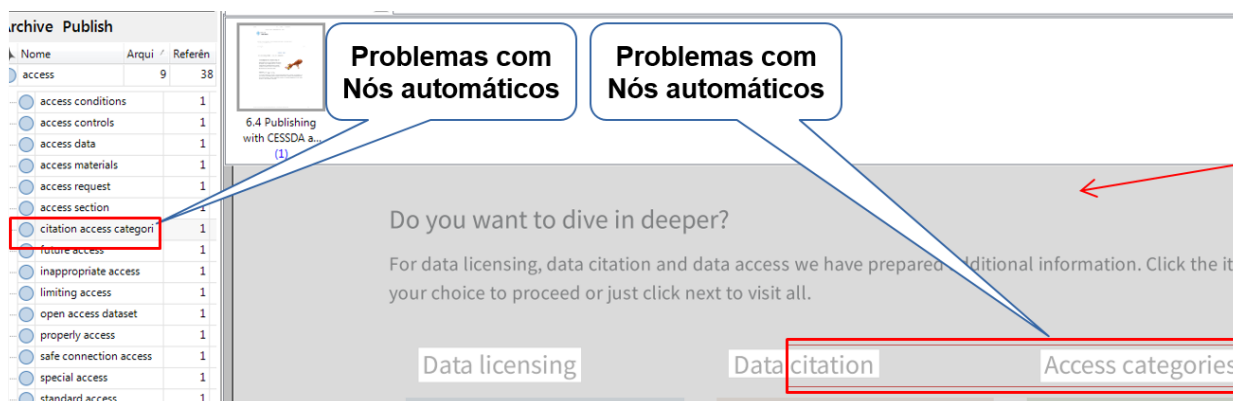
Nome	Arquivos	Referências
sharing	4	7
long-term	3	8
ion	3	8
organizat ion	3	4
preservat ion techniques	3	4
access	3	9
management	4	10
research	4	14
data	4	35

Fonte: Autoria própria com codificação do *software NVivo* (2019).

Outro exemplo de erro de leitura de arquivos convertidos para pdf, utilizando palavras descontextualizadas para formar um termo de possível análise é ilustrado na Figura 31. Neste exemplo, o termo '*citation access categories*' não está agrupado no original, sendo que os termos '*citation*' e '*access categories*' fazem parte de contextos distintos.

Então, optou-se pela codificação manual em substituição à codificação automática como solução para a análise. Formaram-se estratégias de busca para cada requisito e competência previamente identificados nas outras etapas (análises de *frameworks* e de *datasets*). Cada estratégia de busca está listada no Apêndice O e foi executada aí sim com apoio do *Software NVivo* na localização dos termos pesquisados dentro do conteúdo dos documentos que compõem os recursos de aprendizagem.

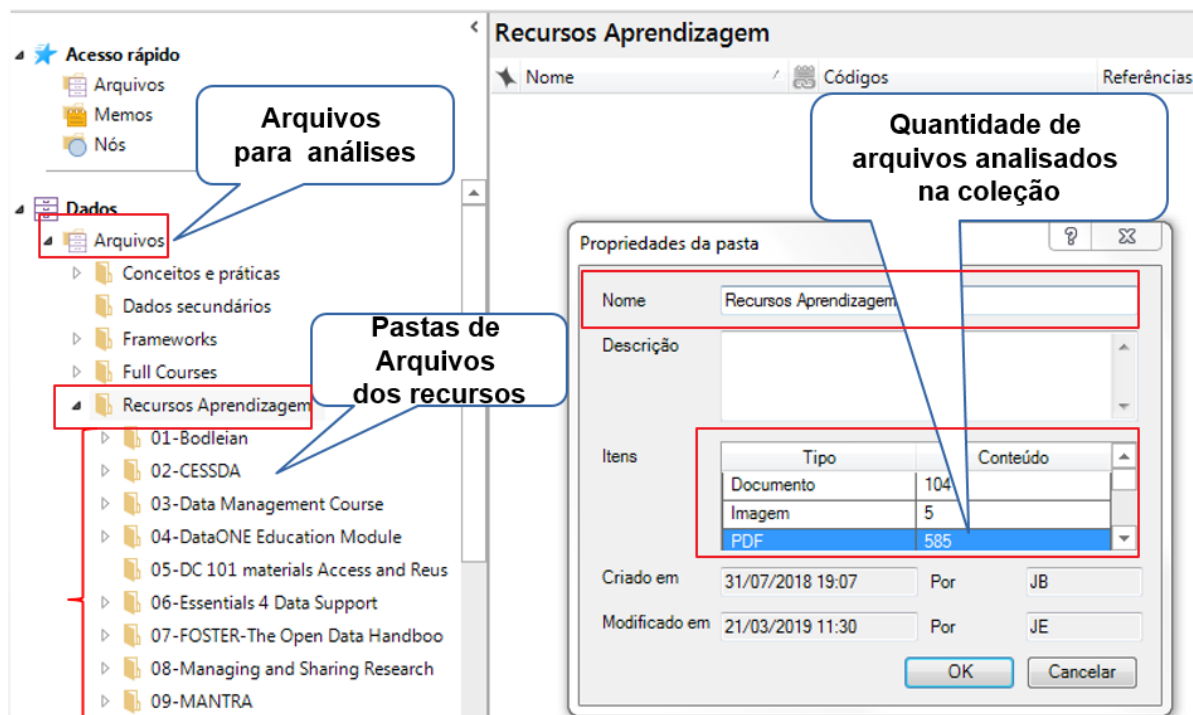
Figura 31 - Problemas com codificação automática no *software NVivo*



Fonte: Autoria própria com codificação do *software NVivo* (2019).

Os arquivos analisados nesta etapa, já armazenados no *Software NVivo*, totalizando 694 arquivos, estão ilustrados na Figura 32. A estrutura do diretório de pastas foi organizada pelo nome dos recursos de aprendizagem *online*.

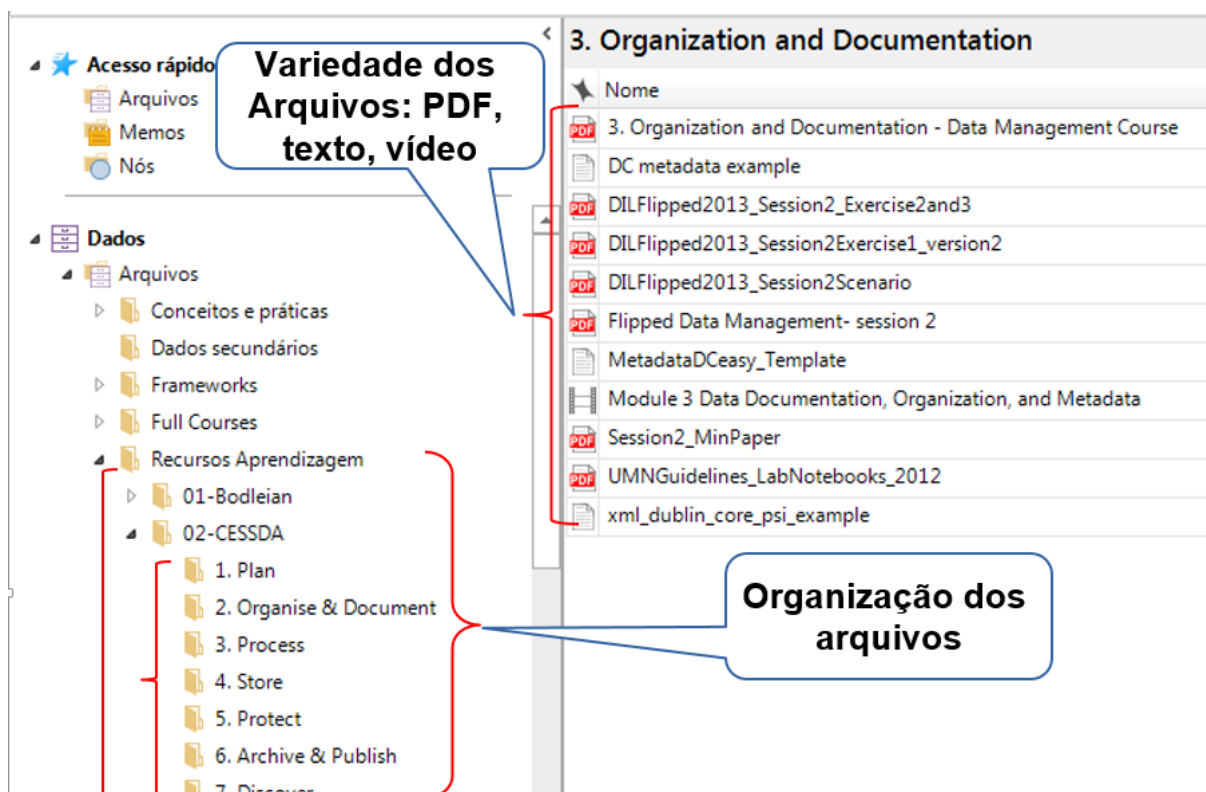
Figura 32 - Conteúdos analisados nos recursos de aprendizagem *online*



Fonte: Autoria própria (2019), com base no Resultado do *Software NVivo* com Recursos de Aprendizagem listados no Quadro 4.

A estrutura do diretório de pastas para acomodar os arquivos dos recursos de aprendizagem, organizados pelos nomes dos recursos e por suas subdivisões de acordo com os temas cobertos em cada menu apresentado nos seus respectivos sítios são exemplificados na Figura 33. Essa figura também apresenta a variedade dos formatos de arquivos analisados.

Figura 33 - Conteúdos analisados nos recursos de aprendizagem online



Fonte: Autoria própria (2019), com base no Resultado do *Software NVivo* com Recursos de Aprendizagem listados no Quadro 4.

A partir dos resultados identificados, iniciou-se a fase (iii) 3 – Tratamentos dos resultados da análise de conteúdo, em que os conjuntos de documentos foram separados de acordo com a quantidade das recorrências de cada requisito e competência. Para os resultados acima de 100 recorrências de uma competência, contida no conjunto completo de todos os documentos (a soma dos recursos A ao K), a estratégia de busca foi replicada para cada subconjunto de arquivos, ou seja, foi refeita em cada um dos recursos de aprendizado separadamente. Assim, obteve-se o desdobramento de cada competência para cada um dos 11 recursos disponíveis, atendendo as seguintes unidades de registro: *15.1 Backup 01.3 Access data, 06.1 Data Management Plan, 16.1 Databases, 16.2 Data collections, 08.4 Licenses types and copyright, 08.2 Citation, 04.2 Archiving practices, 07.2 Documentation Methods, 10.2 Open Access, 06.3 Naming conventions, 08.3 Intellectual Property Rights, 09.2 Metadata Standards*, totalizando 13 competências. Para esses requisitos e suas competências, as buscas foram reproduzidas 143 vezes (uma vez para cada um dos recursos analisados).

Para resultados abaixo de 100 recorrências, a estratégia de busca foi aplicada apenas ao conjunto total dos recursos disponíveis. Isso foi necessário porque notou-se que os resultados da estratégia de busca de algumas competências, quando aplicados em determinados recursos

isoladamente, não recuperavam qualquer resultado. Os requisitos e suas competências que se enquadraram nesse perfil estão descritas na Tabela 15, totalizando 32 competências:

Tabela 15 - Competências nos recursos de aprendizagem com menos de 100 recorrências

Competências	Arquivos	Referências
03.2 Collecting Data	56	89
10.5 Open Data use and reuse	48	70
05.2 Data Format	45	66
10.1 Concept of Open Science	25	54
15.2 Security	40	51
05.4 Data Processing	33	46
12.1 Quality Assurance	35	44
04.1 Plan for archiving	16	39
02.1 Reading	30	38
13.3 Re-use policies	8	36
06.2 Directory structures	19	31
06.4 Organizational planning	24	31
09.1 Metadata Description	23	31
05.1 Data Conversion	16	27
13.1 Reproducible Research	10	23
07.1 Data sets description	18	18
14.1 Re-sharing	8	16
03.1 Obtaining data	9	15
01.2 Discovery	13	14
02.2 Interpreting Understanding data	11	13
08.6 Using data ethically	7	13
12.2 Research methods – practical application	9	13
17.1 Visualization tools	8	12
06.5 Standards	6	7
02.3 Evaluating	6	6
10.3 Open Data Journals	5	6
11.1 Funding policies	6	6
13.2 Legal requirements	4	6
08.1 Apply your licence	1	3
01.1 Access data sources	2	2
04.3 Data Preservation Techniques	2	2
10.4 Open Data Standards	2	2

Fonte: Autoria própria (2019), com base no Resultado do *Software NVivo* com Recursos de Aprendizagem listados no Quadro 4.

Para esses requisitos e suas competências, as buscas foram reproduzidas 32 vezes (uma vez para cada competência em todos os recursos analisados). Com apoio do *Software NVivo* fez-se a seleção das pastas com os arquivos alvos da pesquisa (o conjunto dos recursos ou cada

recurso individualmente) e em seguida a expressão de busca (estratégia de busca das competências) foi executada na caixa de pesquisa. Os resultados foram listados em uma nova aba e puderam ser exportados em formatos de texto, planilha e/ou imagem, como ilustrado na Figura 34.

Figura 34 - Conteúdos analisados nos dos recursos de aprendizagem online

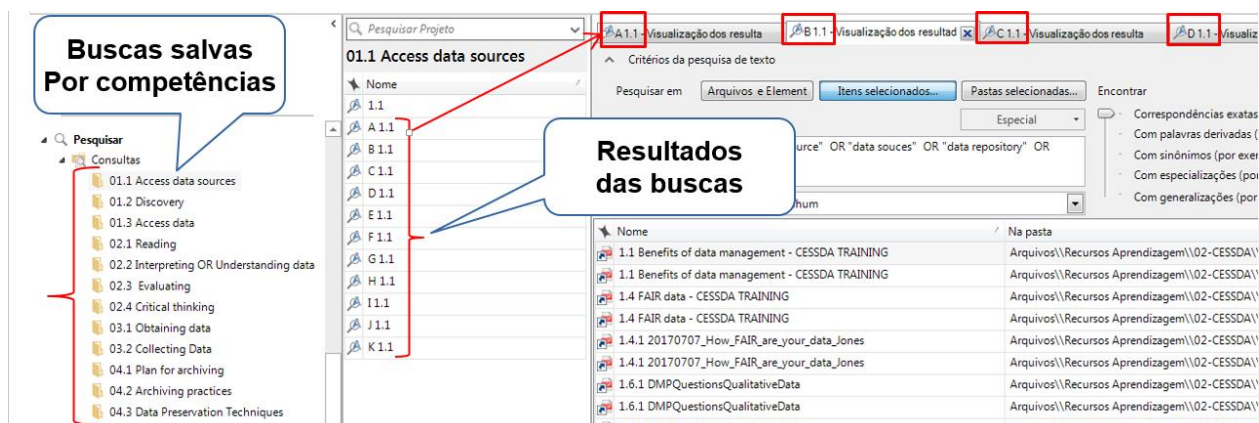
The screenshot shows the NVivo 12 Plus interface. On the left, a project tree lists resources under '01.1 Access data sources', with sub-items A.1.1 through K.1.1. A search window is open, showing search criteria: 'Arquivos e Elementos' and 'Itens selecionados...'. The search query is: "Access data sources" OR "data source" OR "data sources" OR "data repository" OR "access data repository". The search results table is visible below the search window.

Nome	Na pasta	Referências
1.1 Benefits of data management - CESSDA TRAINING	Arquivos\Recursos Aprendizagem\02-CESSDA\1. Plan	1
1.1 Benefits of data management - CESSDA TRAINING	Arquivos\Recursos Aprendizagem\02-CESSDA\2. Orga	1
1.2 Syllabus - Data Management Course	Arquivos\Recursos Aprendizagem\03-Data Managemen	1
1.2.1 Creating a Data Management Plan - University of M	Arquivos\Recursos Aprendizagem\03-Data Managemen	2
1.4 FAIR data - CESSDA TRAINING	Arquivos\Recursos Aprendizagem\02-CESSDA\1. Plan	1
1.4 FAIR data - CESSDA TRAINING	Arquivos\Recursos Aprendizagem\02-CESSDA\2. Orga	1
1.4.1 20170707_How_FAIR_are_your_data_Jones	Arquivos\Recursos Aprendizagem\02-CESSDA\1. Plan	1
1.4.1 20170707_How_FAIR_are_your_data_Jones	Arquivos\Recursos Aprendizagem\02-CESSDA\2. Orga	1
1.6.1 DMPQuestionsQualitativeData	Arquivos\Recursos Aprendizagem\02-CESSDA\1. Plan	9
1.6.1 DMPQuestionsQualitativeData	Arquivos\Recursos Aprendizagem\02-CESSDA\2. Orga	9

Fonte: Autoria própria (2019), com base no Resultado do Software NVivo com Recursos de Aprendizagem listados no Quadro 4.

O critério de salvamento das buscas foi estabelecido pela nomeação de pastas para cada competência pesquisada. As buscas foram replicadas no conjunto de documentos de cada recurso disponível, sendo identificadas pela letra que codifica o recurso (de A até K), acrescida da numeração que identifica cada categoria de análise das competências, que são as unidades de registro. Neste exemplo está a pesquisa da competência 1.1 *Access data sources*, cujo código é o 1.1. A busca foi replicada em cada um dos recursos de aprendizagem, tendo sido gerados resultados nomeados com os códigos A 1.1, B 1.1, C 1.1, D 1.1, E 1.1, F 1.1, G 1.1, H 1.1, I 1.1, J 1.1 e K 1.1, como apresentado na Figura 35. Essa pesquisa foi repetida para todas as unidades de registro, sendo representando neste exemplo a unidade de registro 1.1 *Access data sources*.

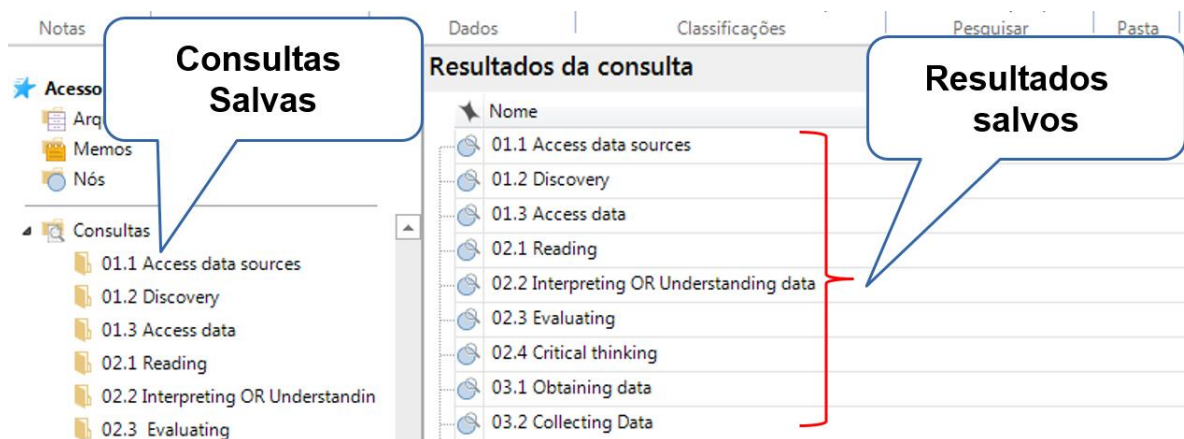
Figura 35 - Critério de salvamento das buscas nos recursos de aprendizagem *online*



Fonte: Autoria própria (2019), com base no Resultado do *Software NVivo* com Recursos de Aprendizagem listados no Quadro 4.

No *Software NVivo* é possível salvar a configuração da busca e dos resultados. Posteriormente, essa busca pode ser repetida com uso de outras fontes para comparação de resultados. A forma de organização e de gravação dos resultados no *Software NVivo* pode ser visualizado na Figura 36.

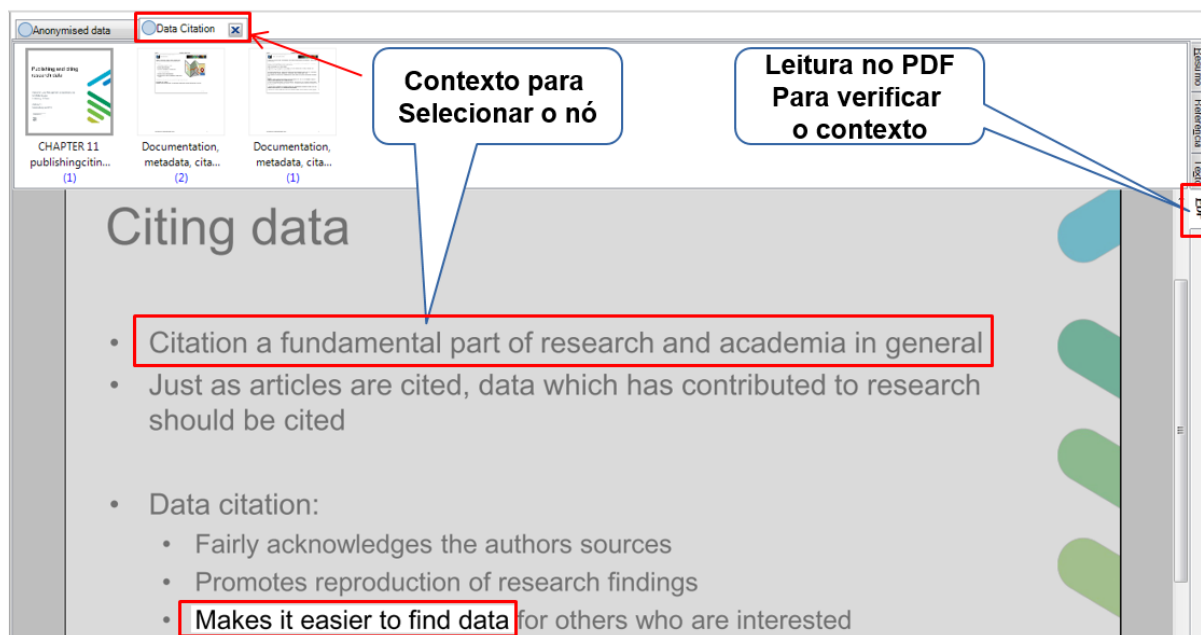
Figura 36 - Resultados salvos nos recursos de aprendizagem *online*



Fonte: Autoria própria (2019), com base no Resultado do *Software NVivo* com Recursos de Aprendizagem listados no Quadro 4.

Os resultados foram exportados em arquivos que relacionam a quantidade de recorrências e os textos que listam a descrição dos textos encontrados, reproduzindo partes do conteúdo dos documentos. A Figura 37 ilustra a como foi realizada a seleção do conteúdo que condiciona o código elaborado para a análise do contexto da referência indicada, sendo selecionado em página específica do documento do recurso.

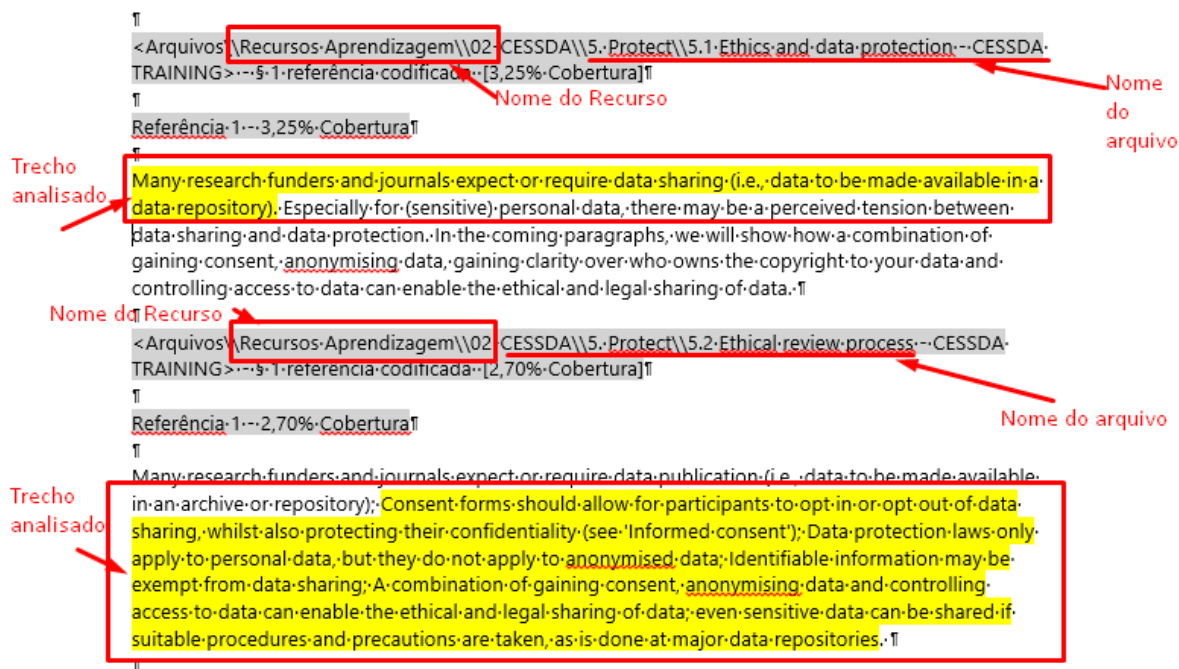
Figura 37 - Seleção do conteúdo para código manual dos recursos de aprendizagem online



Fonte: Autoria própria (2019), com base no Resultado do *Software* NVivo com Recursos de Aprendizagem listados no Quadro 4.

Os resultados apresentados até o momento estão relacionados ao Objetivo Específico 4 – OE4, e foram produzidos a partir das buscas textuais de acordo com as estratégias de pesquisa empregadas para as competências. Esses resultados foram exportados em formato de texto, tendo sido avaliada cada recorrência. O *Software* NVivo permite a extração de trechos localizados na busca, agrupando-os por ordem dos documentos pesquisados. Todos os conteúdos recuperados foram analisados, totalizando 560 páginas com trechos das estratégias de busca executadas. Os trechos selecionados foram traduzidos e, posteriormente, articulado nas descrições para cada competência, apresentados na seção 6.4. Exemplos destas recorrências extraídas estão detalhadas na Figura 38.

Figura 38 - Exemplo de recorrências exportadas para análise



Fonte: Autoria própria a partir das extrações textuais no NVivo (2019).

Os resultados sem recorrências localizados nas buscas foram excluídos da composição do conjunto de requisitos para letramento em dados, sendo eles e suas competências: 02.4 Critical thinking; 05.3 Merging; 05.5 Data Synthesize; 08.5 Publisher rights e 12.3 Research methods – theoretical knowledge, totalizando 5 competências. Para esses requisitos e suas competências, as buscas foram reproduzidas 5 vezes (uma vez para todos os recursos analisados). As estratégias de busca foram repetidas 180 vezes no total. A Tabela 16 lista a quantidade recuperada para cada estratégia de busca nos recursos analisados.

Tabela 16 - Recorrências de todas as competências nos recursos de aprendizagem

(continua)

Competência	Arquivos	Referências
15.1 Backup	126	429
01.3 Access data	205	339
06.1 Data Management Plan	136	310
16.1 Databases	112	304
16.2 Data collections	119	283
08.4 Licenses types and copyright	173	264
08.2 Citation	57	239
04.2 Archiving practices	105	198
07.2 Documentation Methods	97	167
10.2 Open Access	73	154
06.3 Naming conventions	70	148
08.3 Intellectual Property Rights	68	116
09.2 Metadata Standards	53	107
03.2 Collecting Data	56	89
10.5 Open Data use and reuse	48	70
05.2 Data Format	45	66
10.1 Concept of Open Science	25	54
15.2 Security	40	51
05.4 Data Processing	33	46
12.1 Quality Assurance	35	44
04.1 Plan for archiving	16	39
02.1 Reading	30	38
13.3 Re-use policies	8	36
06.2 Directory structures	19	31
06.4 Organizational planning	24	31
09.1 Metadata Description	23	31
05.1 Data Conversion	16	27
13.1 Reproducible Research	10	23
07.1 Data sets description	18	18
14.1 Re-sharing	8	16
03.1 Obtaining data	9	15
01.2 Discovery	13	14
02.2 Interpreting Understanding data	11	13
08.6 Using data ethically	7	13
12.2 Research methods – practical application	9	13
17.1 Visualization tools	8	12
06.5 Standards	6	7
02.3 Evaluating	6	6
10.3 Open Data Journals	5	6
11.1 Funding policies	6	6
13.2 Legal requirements	4	6

(conclusão)

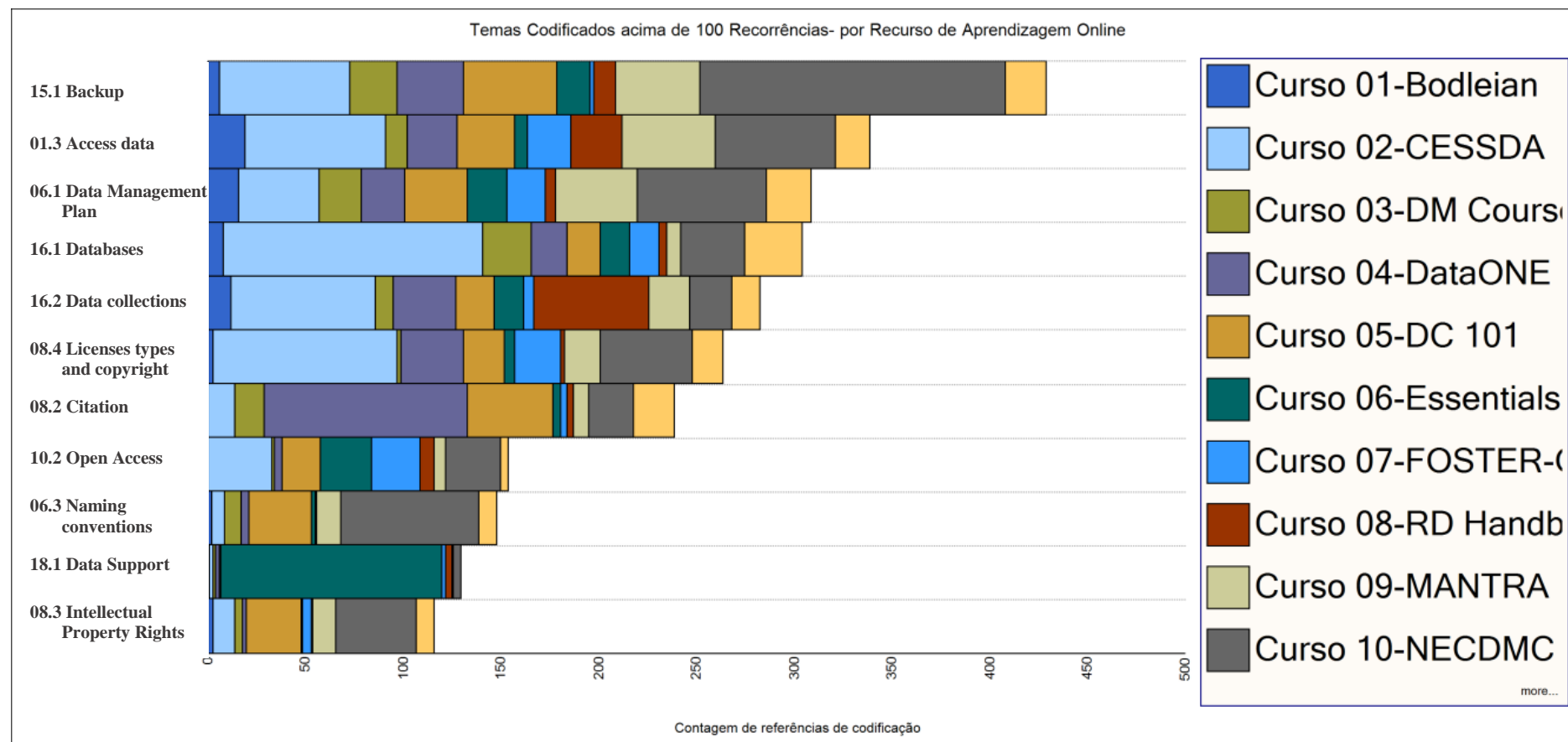
Competência	Arquivos	Referências
08.1 Apply your licence	1	3
01.1 Access data sources	2	2
04.3 Data Preservation Techniques	2	2
10.4 Open Data Standards	2	2
02.4 Critical thinking	0	0
05.3 Merging	0	0
05.5 Data Synthesize	0	0
08.5 Publisher rights	0	0
12.3 Research methods – theoretical knowledge	0	0

Fonte: Autoria própria (2019), com base no Resultado do *Software* NVivo com Recursos de Aprendizagem listados no Quadro 4.

Legenda:	Resultados acima 100 recorrências	Resultados abaixo de 100 recorrências	Buscas sem Resultados
-----------------	--------------------------------------	--	--------------------------

As competências que tiveram mais de 100 recorrências nas buscas estão elencadas no Gráfico 12. Os cursos 02 - CESSDA e o 10-NECDMC são os que apresentam mais conteúdo dessas competências – lembra-se que as finalidades dos Cursos são listadas no Quadro 4. Salienta-se que apenas o recurso 04 - DataOne, apresenta mais recorrência para o requisito 08.2 - *Citation*.

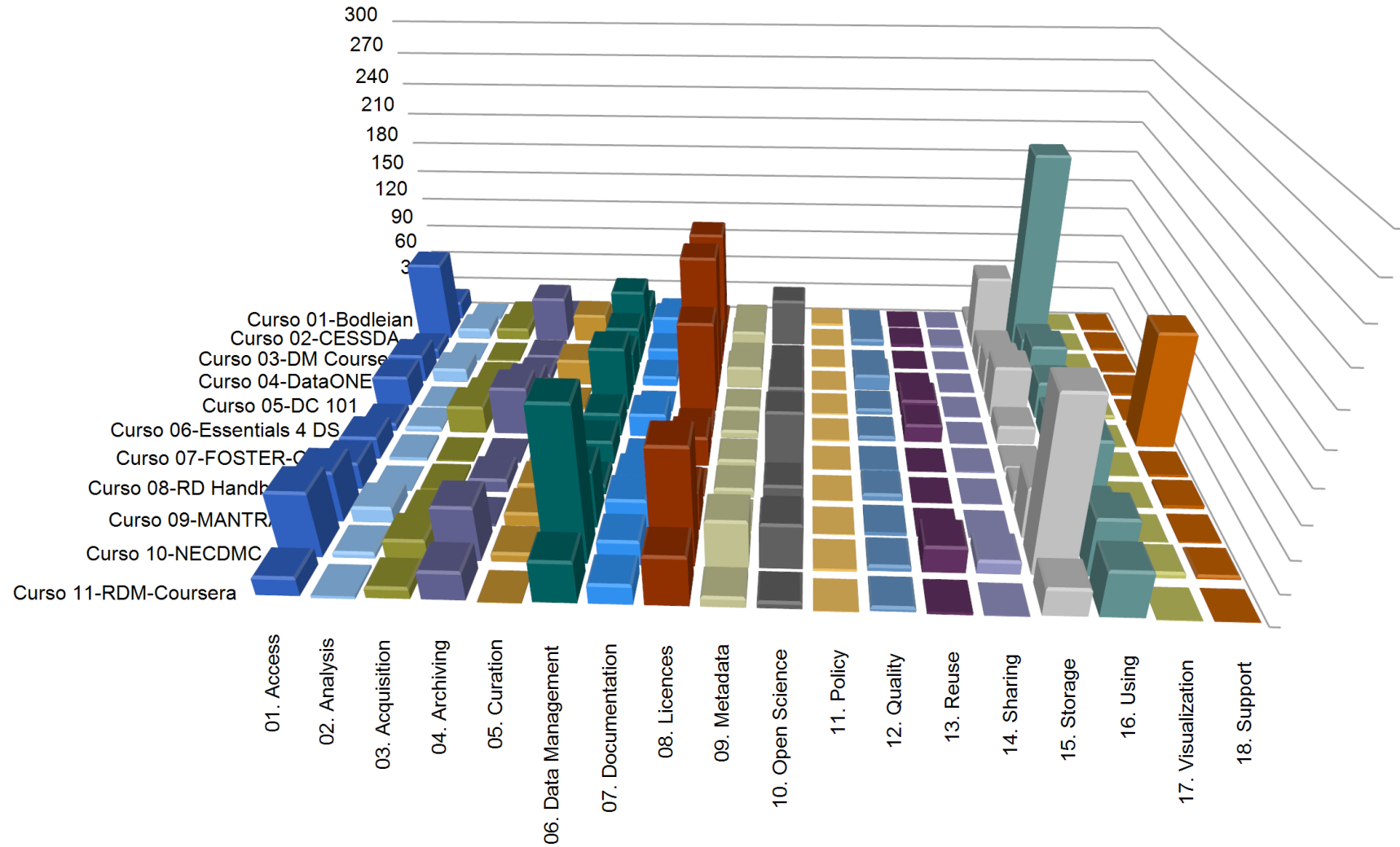
Gráfico 12 - Competências com mais de 100 recorrências - por recurso de aprendizagem



Fonte: Autoria própria (2019), com base no Resultado do *Software NVivo* com Recursos de Aprendizagem listados no Quadro 4.

Observação: a visualização desses requisitos e competências de forma agrupada e distribuída por recursos de aprendizagem está disponível no Gráfico 13. A lista de cada competência e a respectiva quantidade de recorrência está disponível no Apêndice P.

Gráfico 13 - Recursos de Aprendizagem Online por Recorrência de Codificação -



Fonte: Autoria própria (2019), com base no Resultado do Software NVivo com Recursos de Aprendizagem listados no Quadro 4.

Um levantamento por palavras mais recorrentes também foi realizado em todos os recursos de aprendizagem, resultando nas palavras mais frequentes, como apresentado no Quadro 22. Para esse relatório foi utilizado o recurso de agrupamento de palavras similares, o que favorece a acuracidade do resultado, evitando repetições de palavras sinônimas.

Quadro 22 - Frequência de palavras nos Recursos de aprendizagem online

Palavra	Contagem	Palavras similares
data'	8764	data, data', 'data, data#metadata, data'
researchers'	1641	research, 'research, research', researched, researcher, researchers, researchers', researching
managing	1167	manage, manageable, managed, management, management', management', manager, managers, managing
access	1099	access, access', accessed, accesses, accessibility, accessible, accessibly, accessing, accession, accessioned, accessions
backup	872	backup, backups
plans	868	plan, plan', planned, planning, plans
filing	823	file, file', 'file, files, files', files', filing, filings
collects	812	collect, collected, collecting, collection, collections, collects
using	810	used, useful, usefulness, uses, uses', using
archiving	799	archival, archive, archived, archives, archives', archiving, 'archiving
repository	710	repositories, repositories', repository, repository'
documents	650	document, documentation, documented, documenting, documents
metadata	600	metadata, metadata', metadata'
sharing	544	share, shared, shares, sharing
licensing	540	license, licensed, licenses, licensing, licensing', 'licensing
citations	535	citation, citations
projects	520	project, 'project', projects
start	509	start, started, starting, starts
storage	496	storage, 'storage'
formats	490	format, formats, formatted, formatting

Fonte: Autoria própria (2019), com base no Resultado do Software NVivo com Recursos de Aprendizagem listados no Quadro 4.

A partir do resultado descrito no Quadro 22 foi possível gerar uma nuvem de palavras, representando os termos mais recorrentes nos recursos de aprendizagem, como ilustra a Figura 39. As palavras mais recorrentes foram: *data'*; *researchers'*; *managing*; *access*; *backup*; *plans*; *filing*; *collects*; *using*; *archiving*; *repository*; *documents*; *metadata*; *sharing*; *licensing*; *citations*; *projects*; *start*; *storage* e *formats*.

Figura 39 - Nuvem de palavras com 20 termos mais recorrentes nos recursos de aprendizagem



Fonte: Autoria própria (2019), com base no Resultado do *Software NVivo* com Recursos de Aprendizagem listados no Quadro 4.

A partir da análise dos dados apresentados, foi possível segmentar os requisitos identificados inicialmente como os mais adequados ao reúso. Foram resgatadas as competências ligadas apenas aos requisitos comuns para o reúso de dados no contexto desses ambientes, contidos nos quadros conceituais (*frameworks*). Dessa forma, os requisitos e as competências para o reúso foram explorados nos recursos de aprendizagem e serão detalhados na seção a seguir. As competências comuns ao compartilhamento de dados, não aplicadas ao reúso, foram excluídas da análise, sendo elas: *04.1 Plan for archiving*, *04.2 Archiving practices*, *04.3 Data Preservation Techniques*, *06.2 Directory structures*, *06.4 Organizational planning*, *06.5 Standards*, *07.2 Documentation Methods*, *08.1 Apply your licence*, *10.4 Open Data Standards* e *18. Support*. As competências que não tiveram resultados apresentados, também foram excluídas da análise: *02.4 Critical thinking*, *05.3 Merging*, *05.5 Data Synthesize*, *08.5 Publisher rights* e *12.3 Research methods – theoretical knowledge*.

6.2.2 Resultados dos Recursos de Aprendizagem – segmentação requisitos para reúso

Os requisitos identificados para reúso, seguidos pela quantidade de arquivos e de referências (número de codificações) são apresentados no Quadro 23. Pelo critério de quantidade de arquivos localizados, os requisitos *08. Licenses*, *01. Access* e *16. Using* são os

mais representativos dentre os recursos analisados.

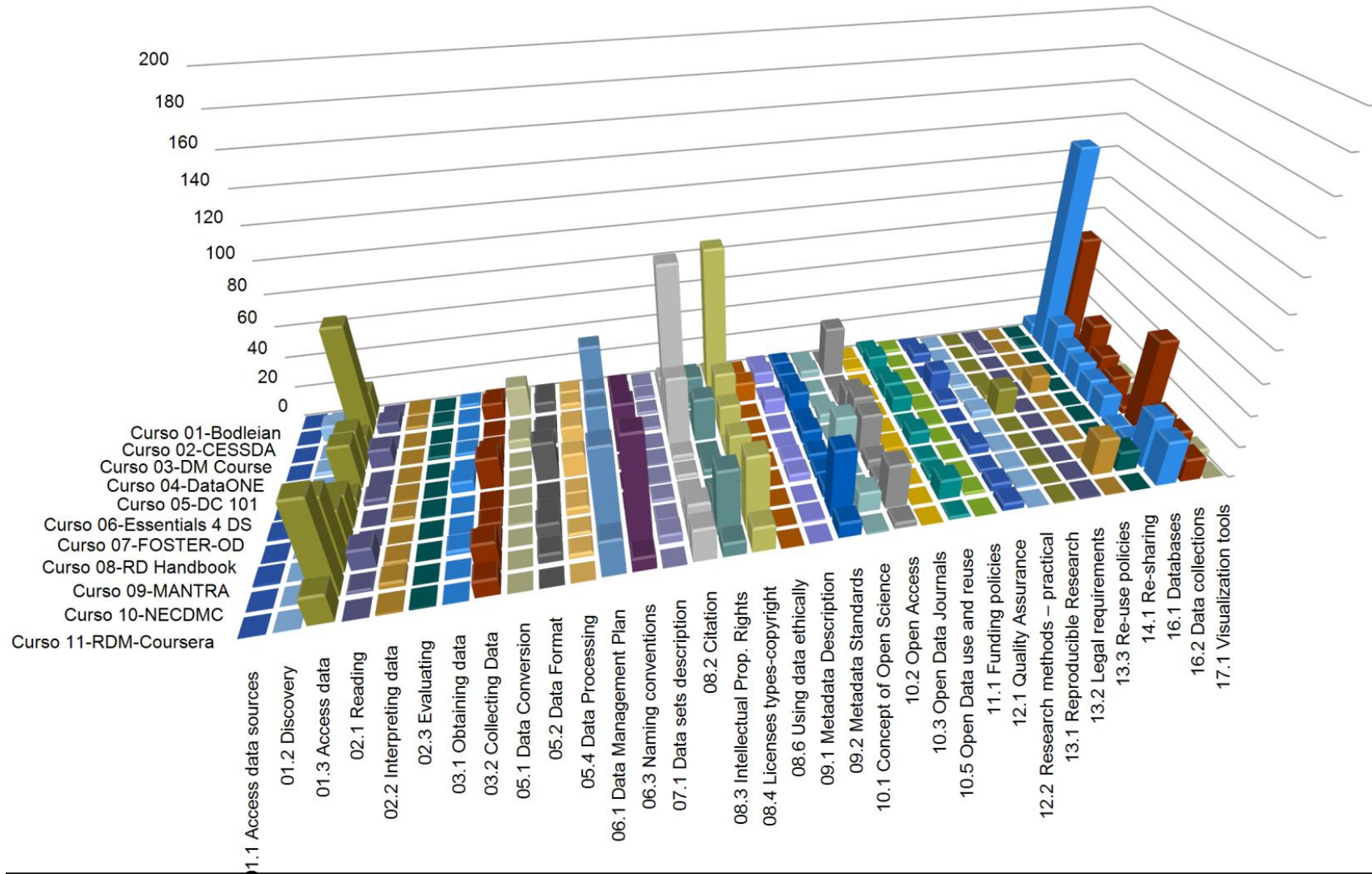
Quadro 23 - Requisitos identificados para reúso - recursos de aprendizagem

Códigos	Número de itens codificados	Número de referências de codificação
01. Access	207	355
02. Analysis	46	57
03. Acquisition	64	104
05. Curation	85	139
06. Data Management	187	527
07. Documentation	106	185
08. Licences	243	635
09. Metadata	64	138
10. Open Science	118	286
12. Quality	43	57
13. Reuse	22	65
14. Sharing	8	16
16. Using	190	587
17. Visualization	8	12

Fonte: Autoria própria (2019), com base no Resultado do *Software NVivo* com Recursos de Aprendizagem listados no Quadro 4.

Sequencialmente identificaram-se as competências ligadas aos requisitos de reúso. A recorrência das competências para reúso, distribuídas pelos recursos de aprendizagem é apresentada no Gráfico 14. São elas: *01.1 Access data sources, 01.2 Discovery, 01.3 Access data, 02.1 Reading, 02.2 Interpreting data, 02.3 Evaluating, 03.1 Obtaining data, 03.2 Collecting Data, 05.1 Data Conversion, 05.2 Data Format, 05.4 Data Processing, 06.1 Data Management Plan, 06.3 Naming conventions, 07.1 Data sets description, 08.2 Citation, 08.3 Intellectual Property Rights, 08.4 Licenses types-copyright, 08.6 Using data ethically, 09.1 Metadata Description, 09.2 Metadata Standards, 10.1 Concept of Open Science, 10.2 Open Access, 10.3 Open Data Journals, 10.5 Open Data use and reuse, 11.1 Funding policies, 12.1 Quality Assurance, 12.2 Research methods – practical, 13.1 Reproducible Research, 13.2 Legal requirements, 13.3 Re-use policies, 14.1 Re-sharing, 16.1 Databases, 16.2 Data collections e 17.1 Visualization tools*. As quantidades listadas por competências representadas neste gráfico 14 estão disponíveis no Apêndice Q.

Gráfico 14 - Recorrência das competências de acordo com os recursos de aprendizagem



Fonte: Autoria própria (2019), com base no Resultado do *Software* NVivo com Recursos de Aprendizagem listados no Quadro 4.

As competências são apresentadas por ordem de relevância dentro dos recursos, como mostra o Quadro 24. As competências para reuso pelo critério de quantidade de referências codificadas acima de 100 recorrências foram: *01.3 Access data*, *16.1 Databases*, *06.1 Data Management Plan*, *16.2 Data collections*, *08.4 Licenses types and copyright*, *08.2 Citation*, *10.2 Open Access*, *06.3 Naming conventions* e *08.3 Intellectual Property Rights*.

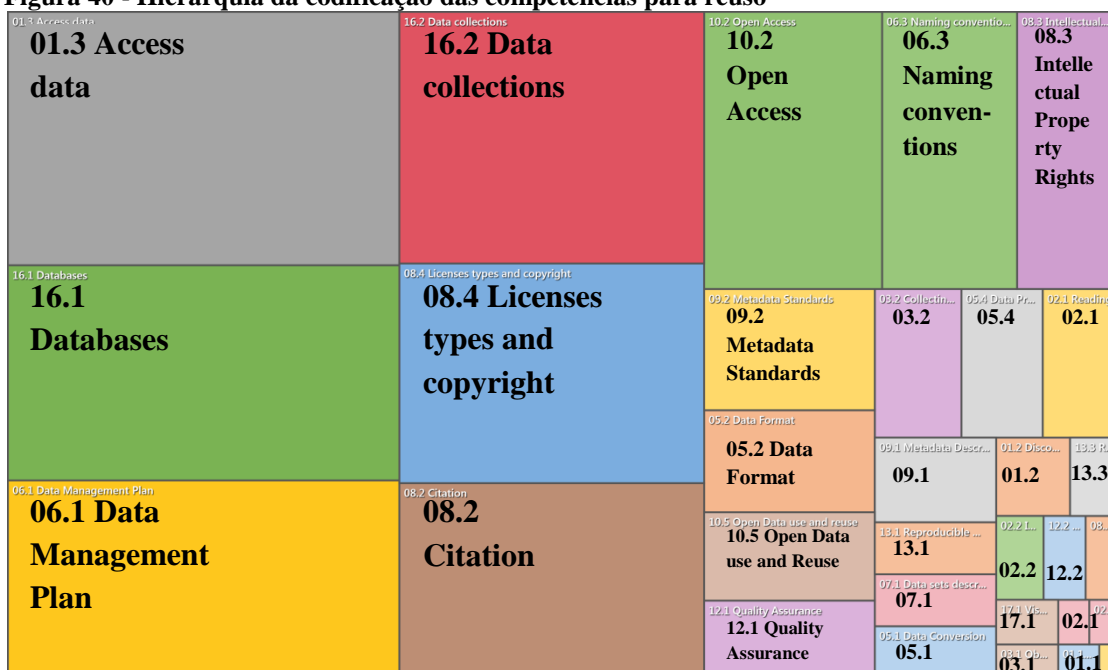
Quadro 24 - OE4: Competências para o reuso de dados – análise dos recursos de aprendizagem

Códigos	Número de referências de codificação	Número de itens codificados
<i>01.3 Access data</i>	425	19
<i>16.1 Databases</i>	362	16
<i>06.1 Data Management Plan</i>	332	14
<i>16.2 Data collections</i>	329	16
<i>08.4 Licenses types and copyright</i>	287	12
<i>08.2 Citation</i>	255	9
<i>10.2 Open Access</i>	211	12
<i>06.3 Naming conventions</i>	161	7
<i>08.3 Intellectual Property Rights</i>	124	7
<i>09.2 Metadata Standards</i>	89	9
<i>05.2 Data Format</i>	75	9
<i>10.5 Open Data use and reuse</i>	65	7
<i>12.1 Quality Assurance</i>	56	8
<i>03.2 Collecting Data</i>	55	5
<i>05.4 Data Processing</i>	52	6
<i>02.1 Reading</i>	50	7
<i>09.1 Metadata Description</i>	44	7
<i>05.1 Data Conversion</i>	27	1
<i>07.1 Data sets description</i>	27	8
<i>13.1 Reproducible Research</i>	27	4
<i>01.2 Discovery</i>	25	3
<i>02.2 Interpreting Understanding data</i>	17	5
<i>13.3 Re-use policies</i>	17	2
<i>12.2 Research methods – practical application</i>	15	2
<i>08.6 Using data ethically</i>	13	1
<i>17.1 Visualization tools</i>	12	1
<i>03.1 Obtaining data</i>	9	1
<i>01.1 Access data sources</i>	6	3
<i>02.3 Evaluating</i>	6	2
<i>10.3 Open Data Journals</i>	6	1
<i>14.1 Re-sharing</i>	3	3

Fonte: Autoria própria (2019), com base no Resultado do Software NVivo com Recursos de Aprendizagem listados no Quadro 4.

A visualização hierárquica dessas competências para o reuso é apresentada na Figura 40:

Figura 40 - Hierarquia da codificação das competências para reuso

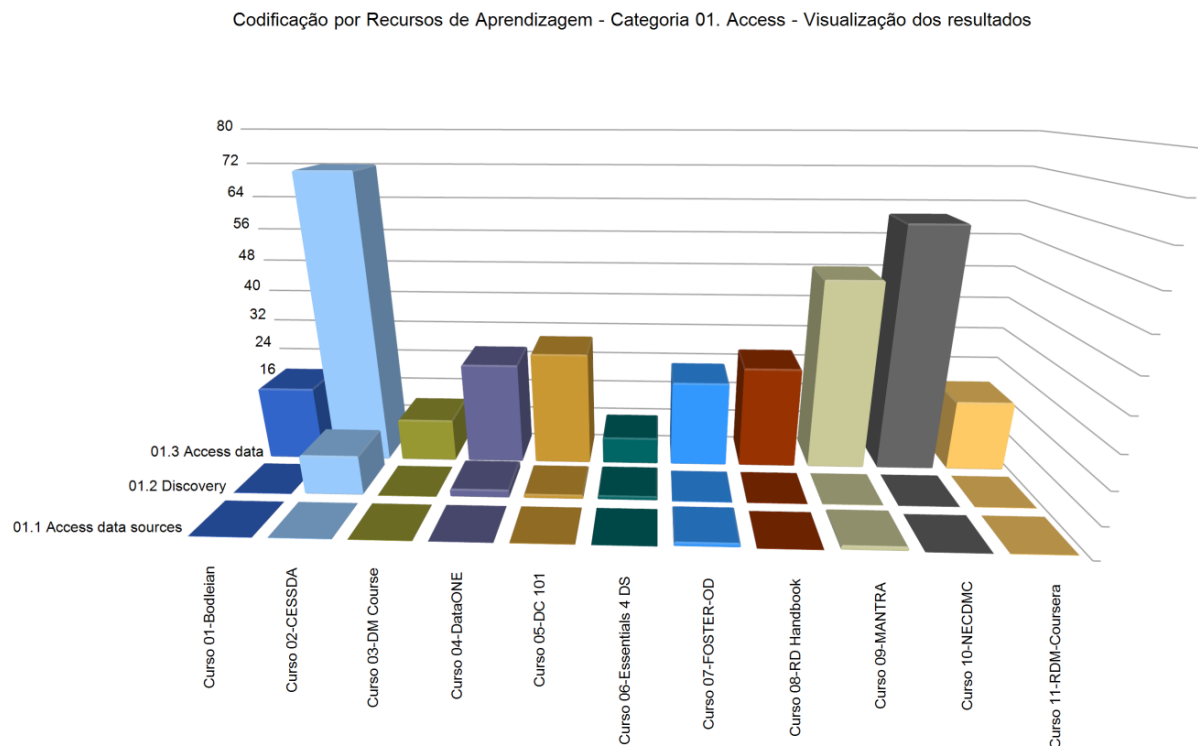


Fonte: Autoria própria (2019), com base no Resultado do *Software NVivo* com Recursos de Aprendizagem listados no Quadro 4.

As competências mais recorrentes nos recursos foram: *01.3 Access data*, *16.1 Databases*, *06.1 Data Management Plan*, *16.2 Data collections*, *08.4 Licenses types and copyright*, *08.2 Citation*, *10.2 Open Access*, *06.3 Naming conventions* e *08.3 Intellectual Property Rights*, e seus resultados são apresentados sequencialmente.

O resultado das recorrências do requisito *1 – Access* e as competências relacionadas, *01.1 Access data sources*, *01.2 Discovery* e *01.3 Access data* são apresentados no Gráfico 15. Neste exemplo, apresentam-se as recorrências dessas competências nos recursos de aprendizagem. A competência *01.3 Access data* tem maior recorrência dentro do requisito *1-Access*, indicando a importância de se ter acesso aos dados. A competência *01.1 Access data sources* não é um termo frequente nos recursos, o que pode indicar que os pesquisadores estejam familiarizados com os repositórios.

Gráfico 15 - Resultado da estratégia de busca do requisito 1 – Access.

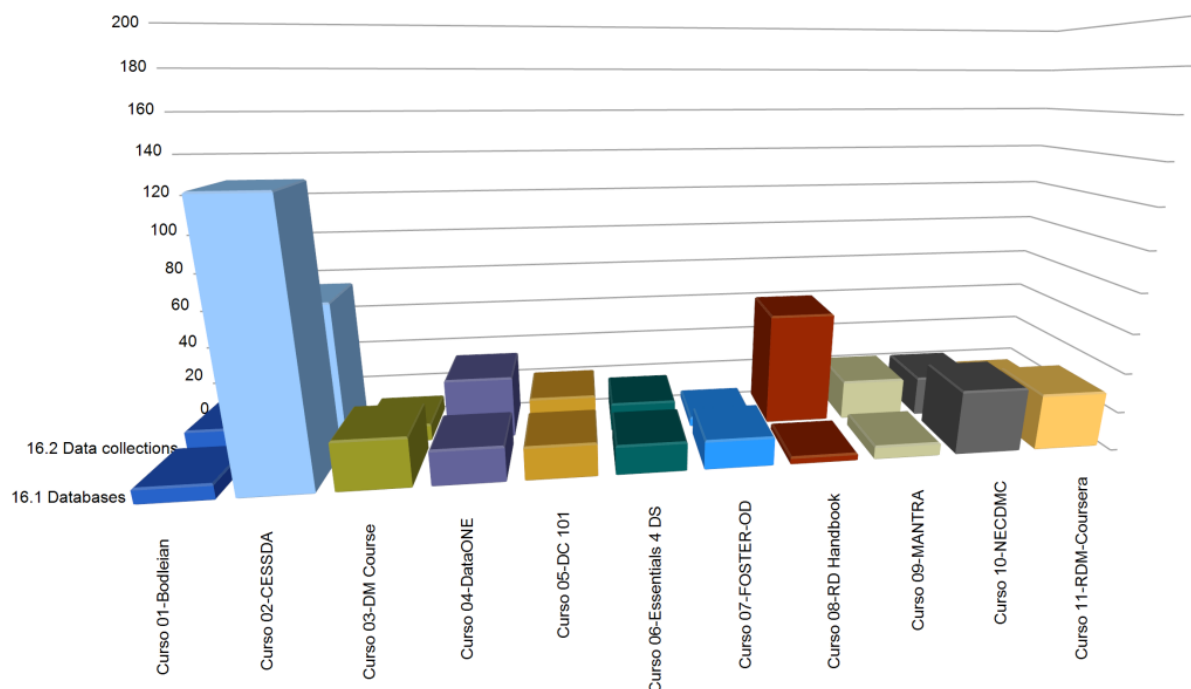


Fonte: Autoria própria (2019)

O resultado da estratégia de busca do requisito 16 - *Using*, com as competências relacionadas com esse contexto: 16.1 *Databases* e 16.2 *Data collections*, é apresentado no Gráfico 16. O curso que mais aborda essas competências é o 02 - CESSDA, indicando a relevância dos conjuntos e bases de dados no reuso de dados em Ciências Sociais, visto este curso ser direcionado para os pesquisadores desta área.

Gráfico 16 - Resultado da estratégia de busca do requisito 16 - Using

Codificação por Recursos de Aprendizagem - Categoria 16. Using - Visualização dos resultados

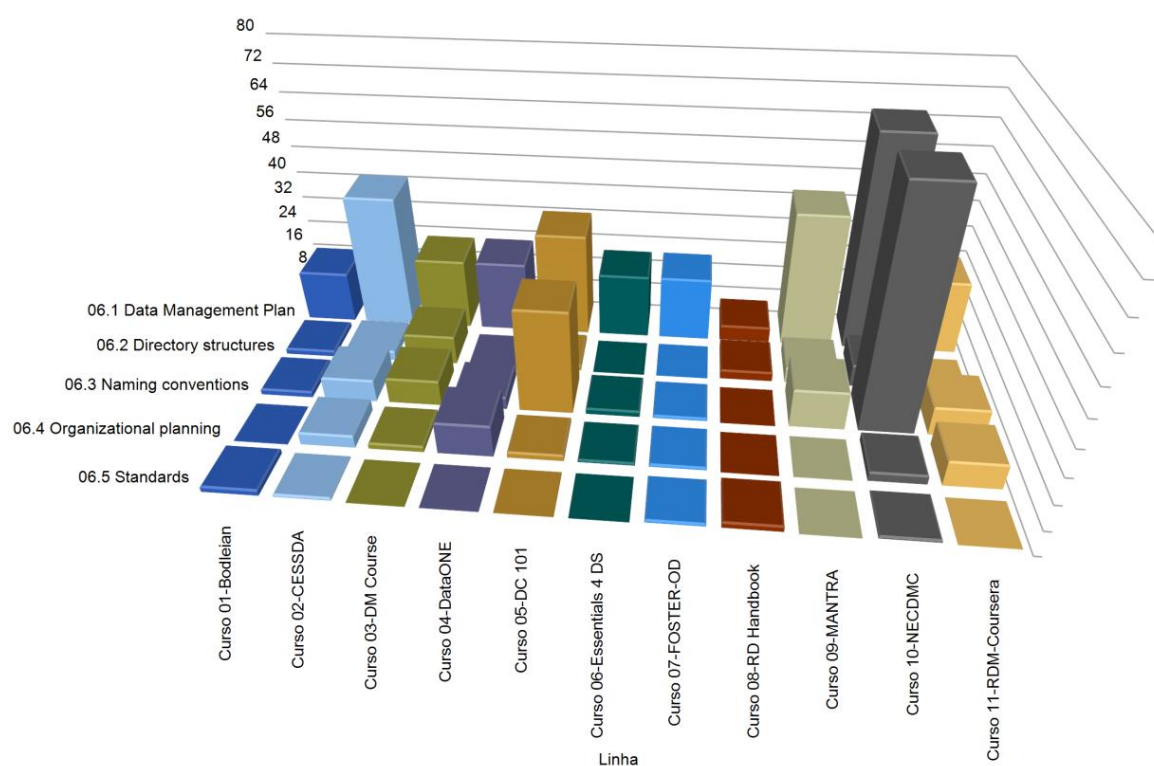


Fonte: Autoria própria (2019)

O resultado da estratégia de busca do requisito 6 – *Data Management*, as competências relacionadas com esse requisito: *06.1 Data Management Plan*; *06.2 Directory structures*; *06.3 Naming conventions*; *06.4 Organizational planning*; *06.5 Standards* são demonstrados no Gráfico 17. As competências *06.1 Data Management Plan* e *06.3 Naming conventions* são as mais expressivas no recurso 10-NECDMC, demonstrando a relevância desses recursos, como os planos de gestão de dados tão exigidos por agências de fomento e por revistas, assim como a importância da correta nomenclatura dos arquivos, tanto para reuso quanto para correta atribuição de metadados conforme corrobora a Revisão de Literatura (CARLSON *et al.*, 2011; CARLSON; JOHNSTON, 2015; GIARLO, 2013; RILEY; TREVOR, 2016; ROYAL SOCIETY, 2012; TENOPIR *et al.*, 2014; WOOD, 2013).

Gráfico 17 - Resultado da estratégia de busca do requisito 6 – Data Management

Codificação por Recursos de Aprendizagem - Categoria 06.Data Management - Visualização dos resultados

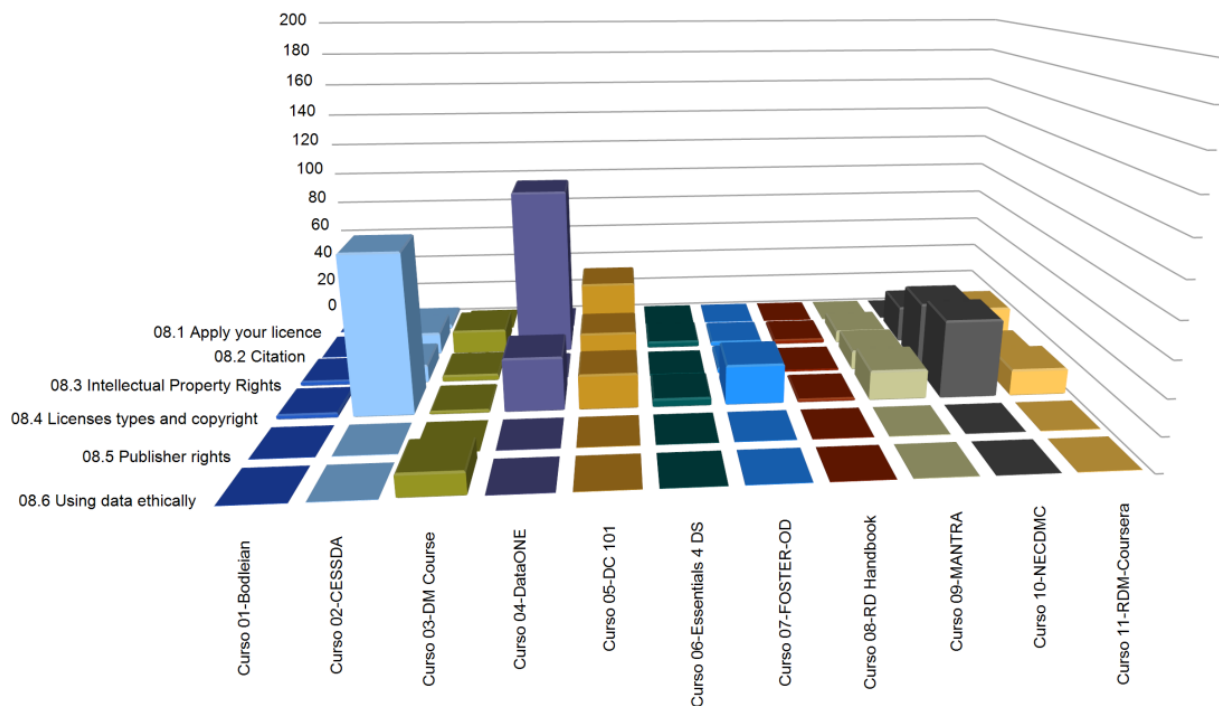
**Fonte: Autoria própria (2019)**

O resultado da estratégia de busca do requisito 08 - *Licenses*, as competências relacionadas com esse contexto: *08.2 Citation*, *08.3 Intellectual Property Rights*, *08.4 Licenses types and copyright* são apresentados no Gráfico 18.

A competência *08.3 Intellectual Property Rights* é mais expressiva no Recurso 02 - CESSDA, enquanto a competência *08.2 Citation* é mais recorrente no recurso 04 - DataOne. Isso demonstra uma forte abordagem na questão de reúso ético, respeitando os direitos autorais e a correta citação dos conjuntos de dados a serem usados, também reforçado pela literatura (ASSOCIATION OF COLLEGE AND RESEARCH LIBRARIES, 2013; CARLSON *et al.*, 2011; CARLSON; JOHNSTON, 2015; FACILITATE OPEN SCIENCE TRAINING FOR EUROPEAN RESEARCH, 2015; TASK GROUP ON DATA CITATION STANDARDS AND PRACTICES, 2013).

Gráfico 18 - Resultado da estratégia de busca do requisito 08 - Licenses

Codificação por Recursos de Aprendizagem - Categoria 08. Licences - Visualização dos resultados

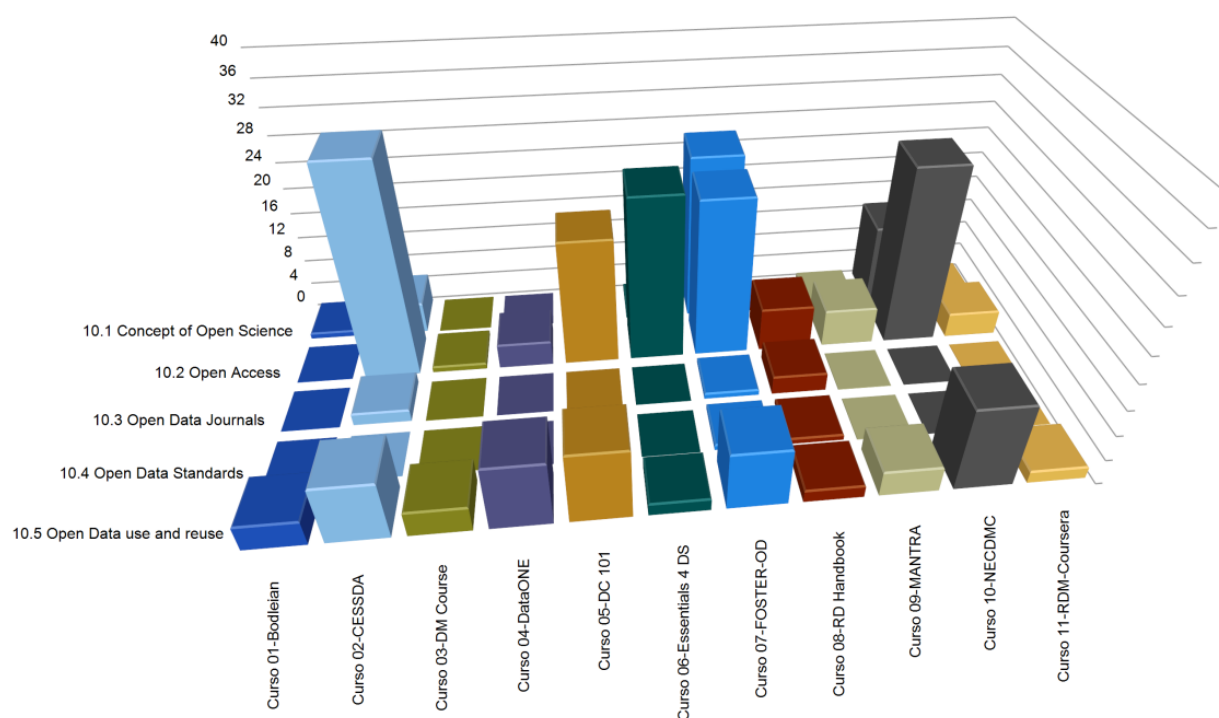


Fonte: Autoria própria (2019)

O resultado da estratégia de busca do requisito 10 - *Open Science*, as competências relacionadas com esse contexto: 10.1 *Concept of Open Science*, 10.2 *Open Access*, 10.3 *Open Data Journal*, 10.4 *Open Data Standards* e 10.3 *Open Data use and reuse* são visualizadas no Gráfico 19. A competência 10.2 *Open Access* é mais abordada no recurso 02-CESSDA, seguida pelos recursos 10 - NECDMC, 07 - Foster, 06 - Essentials 4 Data e DC 101, demonstrando que o acesso aberto é um aspecto central no reuso de dados.

Gráfico 19 - Resultado da estratégia de busca do requisito 10 - Open Science

Codificação por Recursos de Aprendizagem - Categoria 10. Open Science - Visualização dos resultados



Fonte: Autoria própria (2019)

A seção seguinte apresenta o Objetivo Específico 5 – OE5, propondo um conjunto de requisitos com vistas ao reuso de dados nas Ciências Sociais. Esta seção apresenta a comparação em triangulação dos resultados dos Objetivo Específico 2 - OE2, que são os requisitos propostos para o reuso de dados a partir dos quadros conceituais (*frameworks*); com o resultado do Objetivo Específico 3 – OE3, que identificou as perspectivas e as dificuldades dos pesquisadores no reuso dados de pesquisa, por meio de dados secundários (*datasets*); e com o Objetivo Específico 4 – OE4, que identificou as competências essenciais para o reuso de dados no letramento proposto pelos recursos de aprendizado *online*. A próxima seção consolida os resultados das seções 6.1 e 6.2.

6.3 MATRIZES DAS ANÁLISES DE CONTEÚDO

Nesta seção apresenta-se uma matriz com os resultados encontrados nas análises dos *frameworks* (seção 5.9), dos *datasets* (seção 6.1) e dos recursos de aprendizagem *online*, (seção 6.2), buscando elencar os requisitos comuns para o reuso de dados na área das Ciências Sociais e cumprindo assim o Objetivo Específico 5 – OE5. A matriz está organizada por identificação

dos requisitos nas instâncias analisadas, sendo que nos recursos de aprendizagem é possível identificar não apenas a presença de um requisito ou competência, mas também a intensidade destes, por meio da recorrências das estratégias de busca, distribuídas por escala de maiores que 100 recorrências, maiores que 50 recorrências, maiores que 10 recorrências e ainda os que não tiveram recorrências – a identificação consta de legenda de cor junto ao Quadro 25.

As competências não aplicadas ao reuso foram destacadas (**em vermelho**) apenas para identificar a abrangência inicial dos conteúdos analisados, já que os quadros conceituais que as continham foram avaliados na primeira fase. O Quadro 25 apresenta o resultado comparado das análises.

Quadro 25 - OE5: Matriz dos requisitos comuns para reuso de dados – Triangulação

Legenda:	>100 recorrências	>50 recorrências	>10 recorrências	0 recorrências
(continua)				
Matriz dos requisitos comuns para reuso de dados				
Categorias de Análise	Unidades de Registro	Frameworks	Datasets	Recursos de aprendizagem online
1. Access	1.1 Access data sources	✓	✓	✓
	1.2 Discovery	✓	✓	✓
	1.3 Access data	✓	✓	✓
2. Analysis	2.1 Reading	✓	n/a	✓
	2.2 Interpreting/ Understanding data	✓	✓	✓
	2.3 Evaluating	✓	n/a	✓
	2.4 Critical thinking		n/a	n/a
3. Acquisition	3.1 Obtaining data	✓	n/a	✓
	3.2 Collecting Data	✓	✓	✓
4. Archiving	4.1 Plan for archiving	✓	n/a	n/a
	4.2 Archiving practices	✓	n/a	n/a
	4.3 Data Preservation Techniques	✓	n/a	n/a
5. Curation	5.1 Conversion	✓	✓	✓
	5.2 Formats	✓	✓	✓
	5.3 Merging	✓	✓	n/a
	5.4 Processing	✓	✓	✓
	5.5 Synthesize	✓	✓	n/a
6. Data Management	6.1 Data Management Plan	✓	✓	✓
	6.2 Directory structures	✓	n/a	n/a
	6.3 Naming conventions	✓	n/a	✓
	6.4 Organizational planning	✓	n/a	n/a
	6.5 Standards	✓	n/a	n/a

(conclusão)

Categorias de Análise	Unidades de Registro	Matriz dos requisitos comuns para reuso de dados		
		Frameworks	Datasets	Recursos de aprendizagem online
7. Documentation	7.1 Data sets description	✓	✓	✓
	7.2 Documentation Methods	✓	✓	n/a
8. Licences	8.1 Apply your licence	✓	✓	n/a
	8.2 Citation	✓	✓	✓
	8.3 Intellectual Property Rights	✓	✓	✓
	8.4 Licenses types and copyright	✓	✓	✓
	8.5 Publisher rights	✓	✓	n/a
	8.6 Using data ethically	✓	✓	✓
	9. Metadata	9.1 Metadata Description	✓	✓
9.2 Metadata Standards		✓	✓	✓
10. Open Science	10.1 Concept of Open Science	✓	✓	n/a
	10.2 Open Access	✓	✓	✓
	10.3 Open Data Journals	✓	✓	✓
	10.4 Open Data Standards	✓	✓	n/a
	10.5 Open Data use and reuse	✓	✓	✓
11. Policy	11.1 Funding policies	✓	✓	✓
12. Quality	12.1 Quality Assurance	✓	✓	✓
	12.2 Research methods – practical application	✓	n/a	✓
	12.3 Research methods – theoretical knowledge	✓	n/a	n/a
13. Reuse	13.1 Reproducible Research	✓	✓	✓
	13.2 Legal requirements	✓	✓	n/a
	13.3 Re-use policies	✓	✓	✓
14. Sharing	14.1 Re-sharing	✓	n/a	✓
15. Storage	15.1 Backup	✓	n/a	n/a
	15.2 Security	✓	n/a	n/a
16. Using	16.1 Databases	✓	n/a	✓
	16.2 Data collections	✓	✓	✓
17. Visualization	17.1 Visualization tools	✓	n/a	✓
18. Support and formal training	18.1 Support and formal training	n/a	✓	✓

Fonte: Autoria própria (2019).

Desta forma, tem-se a distribuição de todos os requisitos e competências por instância analisada, o que demonstra quais são os requisitos mais relevantes de acordo com essas análises, sendo a base para a construção da proposta desta pesquisa. Após a execução da etapa de coleta e análise dos dados descritos nesta seção, são apresentados os resultados a seguir: a proposta de conjunto de Requisitos e competências para o letramento em reuso de dados, bem como a validação desses elementos em relação as recomendações das melhores práticas W3C.

6.4 PROPOSTA DE UM CONJUNTO DE REQUISITOS PARA LETRAMENTO EM DADOS

A construção dessa proposta está baseada nas necessidades de letramento identificadas na literatura, como os trabalhos de Calzada Prado e Marzal, (2013); Carlson *et al.*(2011); Carlson e Johnston (2015); Mandinach e Gummer (2013); Schneider (2013) e Tenopir *et al.* (2017), apoiados por princípios de instituições das bibliotecas acadêmicas⁵⁰, sendo orientada pelos conceitos da Teoria Ator-Rede. Inferindo que a comunicação dos resultados científicos é um processo sociotécnico, de acordo com Borgman (2007) e Cutcliffe (2003), a proposta considerou uma composição simetricamente generalizada, com o mesmo valor para os atores humanos e não humanos, pela integração dos resultados de conceitos da literatura (*frameworks*), dos atores envolvidos (análise dos *datasets*) e pelos ambientes virtuais de aprendizagem - AVAs (recursos de aprendizagem *online*) a partir dos estudos de Callon(1986) e Latour (1992).

Sob a perspectiva da TAR, os *frameworks* representam as inscrições, apropriando-se da afirmação de Callon, Law e Rip (1986), em que o texto também é um elemento que demonstra como os seus autores constroem as suas versões de mundo. Os *datasets* expressam, em relação aos dados, a realidade das atividades práticas dos cientistas, que Callon, Law e Rip (1986) consideram tão importantes quanto as suas inscrições. Portanto, usou-se os *datasets* como forma de seguir os cientistas, o que Latour (1998; 2000; 2001) demonstra ser a prática que permite descrever o que eles fazem (LATOUR, 2001) ou ainda, neste caso, aplicado ao reuso de dados, de acordo com Curty (2016) permite conhecer como eles se comportam. Ao analisar os Recursos de aprendizagem *online*, destaca-se o papel relevante dos AVAs como condutor da mediação técnica, conforme proposto por Latour (2001). Os AVAs são, na nomenclatura de Latour (2001), os actantes não humanos com os quais os pesquisadores trocam propriedades,

⁵⁰ ACRL Information Literacy Best Practices Committee (2012); Association of College And Research Libraries (2012a); Association of College And Research Libraries (2013).

pois estes interferem na ação humana, atuando como mediadores do aprendizado (papel de um instrutor que não está ali presencialmente), mas está utilizando propriedades técnicas para comunicar a instrução e trasladando o objetivo do cientista ou pesquisador, que é o letramento para o reúso de dados. Apropriando-se da perspectiva de mediação técnica de Latour (2001), os AVAs ensinam aos pesquisadores, trocando propriedades do legado de seus idealizadores. Esta pode ser considerada uma interação heterogênea entre humanos, inscrições, *softwares* e máquinas. São essas associações que para Latour (2006) que conferem a estabilidade dessas entidades, promovendo a sua durabilidade.

Seguindo esta perspectiva da TAR, a partir da análise de conteúdo, foram selecionados 16 categorias de análises e 37 unidades de registro para compor o conjunto de requisitos e competências essenciais ao letramento em dados. A análise de conteúdo considerou os *frameworks*, os *datasets* e os Recursos de Aprendizagem *Online*. Os requisitos comuns para o reúso de dados a partir desta análise correspondem ao objetivo geral (OG), que está proposto no Quadro 26.

Este conjunto de requisitos e competências pode antecipar aos pesquisadores a variedade de desafios existentes durante a reutilização de dados. Pois, para aqueles sem experiência nesta prática, as dificuldades a serem enfrentadas na reutilização podem não ser conhecidas antecipadamente ao reúso efetivo dos dados (CARLSON *et al.*, 2011; CALZADA PRADO; MARZAL, 2013; CARLSON; JOHNSTON, 2015; JEFFRYES; JOHNSTON, 2013; FOSTER, 2015; VITAE, 2011). Na sequência são apresentadas as explicações de cada requisito, de acordo com a estrutura conceitual adotada nesta pesquisa: Categorias de Análise (CA), Unidades de Registro (UR) e Unidades de Contexto (UC).

Quadro 26 - OG: Requisitos e competências para o letramento em reuso de dados

Categorias de Análise	Unidades de Registro	Categorias de Análise	Unidades de Registro
<i>1. Access</i>	<i>1.1 Access data sources</i>	<i>7. Licences</i>	<i>7.3 Licenses types and copyright</i>
	<i>1.2 Discovery</i>		<i>7.4 Using data ethically</i>
	<i>1.3 Access data</i>	<i>8. Metadata</i>	<i>8.1 Metadata Description</i>
<i>2. Analysis</i>	<i>2.1 Reading</i>		<i>8.2 Metadata Standards</i>
	<i>2.2 Interpreting/ Understanding data</i>	<i>9. Open Science</i>	<i>9.1 Concept of Open Science</i>
	<i>2.3 Evaluating / 2.4 Critical thinking</i>		<i>9.2 Open Access</i>
<i>3. Acquisition</i>	<i>3.1 Obtaining data</i>		<i>9.3 Open Data Journals</i>
	<i>3.2 Collecting Data</i>	<i>9.5 Open Data use and reuse</i>	
<i>4. Curation</i>	<i>4.1 Conversion</i>	<i>10. Policy</i>	<i>10.1 Funding policies</i>
	<i>4.2 Formats</i>	<i>11. Quality</i>	<i>11.1 Quality Assurance</i>
	<i>4.3 Merging</i>	<i>12. Reuse</i>	<i>12.1 Reproducible Research</i>
	<i>4.4 Processing</i>		<i>12.2 Legal requirements</i>
	<i>5.5 Synthesize4</i>		<i>12.3 Re-use policies</i>
<i>5. Data Management</i>	<i>6.1 Data Management Plan</i>	<i>13. Sharing</i>	<i>13.1 Re-sharing</i>
	<i>6.2 Naming conventions5</i>	<i>14. Using</i>	<i>14.1 Databases</i>
<i>6. Documentation</i>	<i>6.1 Data sets description</i>		<i>14.2 Data collections</i>
<i>7. Licences</i>	<i>7.1 Citation</i>	<i>15. Visualization</i>	<i>15.1 Visualization tools</i>
	<i>7.2 Intellectual Property Rights</i>	<i>16. Support and formal training</i>	<i>16.1 Support and formal training</i>

Fonte: Autoria própria (2019)

Os requisitos (Categorias de Análise) e as competências essenciais (Unidades de Registro) foram descritas de forma hierárquica, com níveis de recuo, para facilitar a visualização, conforme o Quadro 27. A hierarquia de primeiro nível é a **Categoria de Análise**, representada pelo indicador CA e destacada pela fonte em negrito. A hierarquia de segundo nível é a *Unidade de Registro*, representada pelo indicador UR e destacada em itálico. A hierarquia de terceiro nível é a Unidades de Contexto, representada pelo indicador UC e sem destaque.

As Unidades de Contexto foram identificadas a partir dos *Frameworks* analisados de Calzada Prado e Marzal (2013), Carlson *et al.* (2011), Carlson e Johnston (2015), Foster (2015), Jeffreyes e Johnston (2013) e Vitae (2011). As Categorias de Análise selecionadas dos respectivos *frameworks*, assim como as nomenclaturas desconsideradas, estão detalhados na seção 5.9, especificadas no Quadro 18, cujo detalhamento completo está disponível no Apêndice J. Na estrutura de descrição das competências, foi inserida a indicação numérica referente a cada plano conceitual de origem da unidade de contexto (*Framework* 01 - F01,

Framework 02 - F02, Framework 03 - F03, Framework 04 - F04, Framework 05 - F05 e Framework 06 – F06). O requisito 16. *Support and formal training* só foi encontrado nas análises dos *datasets*, portanto, estão indicados como provenientes do *Dataset 1, Dataset 2 e Dataset 3*. Seguindo a abordagem metodológica de Latour (2000) sobre a importância das descrições, os requisitos foram descritos a partir da análise dos conteúdos dos recursos de aprendizagem *online* (R01, R02, R03, R04, R05, R06, R07, R08, R09, R10 ou R11). As notas de rodapé foram inseridas para permitir a localização do arquivo específico do recurso de aprendizagem que contribuiu com a descrição.

Quadro 27 - Indicador da estrutura dos requisitos e das competências essenciais

Indicador da estrutura	Nome da estrutura	Nível hierárquico
CA	Categorias de Análise	Nível 1
UR	<i>Unidades de Registro</i>	<i>Nível 2</i>
UC	Unidades de Contexto	Nível 3

Fonte: Autoria própria (2019).

O conjunto de requisitos e competências essenciais é apresentado a seguir, indo de Acesso – Requisito 1 até Treinamento e Suporte – Requisito 16:

- Requisito 1 – **Acesso** (*Access*)

O primeiro requisito é o **Acesso** (*Access*). Como pesquisar e acessar os dados são os primeiros passos para reutilização de dados, a disponibilidade de repositórios de dados é fundamental para a localização e o acesso. Existem milhares de repositórios de dados em todo o mundo, abrigando conjuntos de dados de uma ampla variedade de áreas de pesquisa (R02⁵¹). Alguns repositórios mantêm dados específicos do assunto, enquanto outros contêm dados de várias disciplinas diferentes (R07⁵²). O papel dos repositórios é essencial no compartilhamento e no reuso de dados e a sua disponibilidade é um importante fator chave na reutilização de dados

⁵¹ 02-CESSDA\1. Plan\1.6.1 DMPQuestionsQualitativeData

⁵² 07-FOSTER-The Open Data Handbook\07-Full\fooster-Data sources _ Find, reuse and cite data _ Library _ University of Leeds

(R09⁵³) (R02⁵⁴), visto que viabilizam o acesso aos dados (R07⁵⁵) (R02⁵⁶) (R02⁵⁷) (R11⁵⁸). O re3data (re3data.org), Registro de Repositório de Pesquisa de Dados de pesquisa, lista atualmente mais de 2.400 repositórios disciplinares, das mais relevantes universidades e institutos de pesquisa.

Pesquisadores desejam acessar dados de todos os tipos de estudos, metodologias e disciplinas (R01⁵⁹). Entretanto, o acesso às fontes de dados pode ser aberto (sem restrições) ou controlado, por meio de senhas e possíveis embargos, ou ainda limitados aos participantes de um projeto (R01⁶⁰). O pesquisador precisará reconhecer a fonte adequada, confiável, de dados disciplinares ou multidisciplinares (R02⁶¹) (R07⁶²) (R07⁶³). Também será necessário citá-la após o uso dos dados, respeitando a propriedade intelectual da fonte de dados e dos conjuntos de dados (R09⁶⁴). O Plano de gerenciamento de dados deve informar as condições deste acesso, inclusive se haverá categorias e restrições, por meio de acesso controlado e seguro, principalmente aos dados confidenciais (R09⁶⁵). Ele precisará verificar se há consentimento dos participantes e/ou se os dados são anonimizados, e se há política de confidencialidade (R02⁶⁶). A forma de indicação para este requisito é:

CA 1. Acesso. (*Access*)

UR 1.1 *Acessar Fonte de Dados (Access Data Sources)*

UC Localizar fontes de informação adequadas a essa necessidade; Capacidade de acessar fontes de dados apropriadas às informações necessárias (F02). Estar ciente das possíveis fontes de dados, ser capaz de avaliá-las. (F02). Usar diferentes portais de pesquisa do Acesso Aberto (F03). Diferenciar nível de direitos de acesso (F04). Avaliar a confiabilidade, reputação, moeda, autoridade e relevância das fontes. (F04).

⁵³ 09-MANTRA\8 Sharing, preservation & licensing\Sharing, preservation, and licensing15

⁵⁴ 02-CESSDA\7. Discover\7. Discover - CESSDA TRAINING

⁵⁵ 07-FOSTER-The Open Data Handbook\The Open Data Handbook\The Open Data Handbook-How to Open up Data

⁵⁶ 02-CESSDA\5. Protect\5.7 Copyright - CESSDA TRAINING

⁵⁷ 10-NECDMC\7 Repositories, Archiving E Preservation\necdmc_module7

⁵⁸ 11-Research Data Management and Sharing - Coursera\1-Understanding Research Data\1-Understanding Data Management

⁵⁹ 01-Bodleian\Kit-session5-DataSharing\Session5GroupExercise1

⁶⁰ 01-Bodleian\Kit-session5-DataSharing\Session5GroupExercise1

⁶¹ 02-CESSDA\6. Archive & Publish\6.3 Data publishing routes - CESSDA TRAINING

⁶² 07-FOSTER-The Open Data Handbook\07-Full\fooster-Data Protection & Ethics

⁶³ 07-FOSTER-The Open Data Handbook\07-Full\fooster _ Managing and Sharing Research Data

⁶⁴ 09-MANTRA\5 Documentation, metadata, citation\Documentation, metadata, citation20

⁶⁵ 09-MANTRA\2 Data management plans\Data management plans11

⁶⁶ 02-CESSDA\1. Plan\1.3 Data in the social sciences - CESSDA TRAINING

UR 1.2 Descobrir (Discovery)

UC Localizar e utilizar repositórios de dados disciplinares. (F01). Definir com precisão a necessidade informacional (F02). Identificar e acessar recursos bibliográficos apropriados, arquivos e outras fontes de informações relevantes, incluindo recursos baseados na *Web*, fontes primárias e repositórios. (F04)

UR 1.3 Acessar os Dados (Access Data)

UC Identificar os possíveis controles de acesso. Planejar os controles de acesso apropriados. (F06). Relatar o acesso potencial (F06).

- Requisito 2 - **Análise** (*Analysis*).

O segundo requisito é a **Análise** (*Analysis*). Os dados devem estar documentados para que sejam compreendidos por pesquisadores (R02⁶⁷), evitando as limitações da ambiguidade e da semântica das palavras em diferentes disciplinas, minimizando a má interpretação do conteúdo devido à informação fragmentada ou explicando uma métrica utilizada, permitindo o reuso correto (R07⁶⁸). A correta atribuição de metadados aumenta o potencial de reutilização dos dados. Compreender a terminologia de gerenciamento de dados, o ambiente para entender como interpretá-los, que situados, ter acesso à documentação para entender os contextos dos dados (R07⁶⁹). A forma de indicação para este requisito é:

CA 2. Análise (*Analysis*)

UR 2.1 Leitura (Reading)

UC Compreender as várias formas nas quais os dados podem ser apresentados (escritos, numéricos ou gráficos). (F02); Compreender o que se entende por dados (F02).

UR 2.2 Interpretação/ Compreensão dos Dados (Interpreting/ Understanding Data)

UC Compreender as várias formas em que os dados podem ser apresentados (escritos, numéricos ou gráficos), e suas respectivas convenções, e poder interpretá-los. Entender as maneiras de apresentar e representar dados. (F02).

⁶⁷ 02-CESSDA\1. Plan\1.6.1 DMPQuestionsQualitativeData

⁶⁸ 07-FOSTER-The Open Data Handbook\07-Full\A FAIRy tale

⁶⁹ 07-FOSTER-The Open Data Handbook\07-Full\A FAIRy tale

UR 2.3 Avaliação (*Evaluating*)

UC Avaliar os dados criticamente. (F02). Seguir critérios de avaliação de dados (incluindo autoria, método de obtenção e análise de dados, comparação, inferências e resumos de dados). (F02). Analisar as ferramentas necessárias. (F02). Utilizar as ferramentas de gerenciamento de fluxo de trabalho apropriadas para automatizar a análise repetitiva dos dados. (F01).

UR 2.4 Pensamento Crítico (*Critical Thinking*)

UC Analisar e avaliar criticamente os próprios dados e os de outros (F04).

- Requisito 3 – **Aquisição** (*Acquisition*)

O terceiro requisito é a **Aquisição** (*Acquisition*). Algumas barreiras podem ser enfrentadas no momento da aquisição dos dados e uma avaliação também é necessária nesta etapa. Permissões extras para o efetivo *download* dos *datasets*, realização de cadastros, dados sob período de embargo, são algumas delas (R02⁷⁰) (R04⁷¹). Em alguns casos é preciso o consentimento informado no momento de obtenção dos dados de pessoas, respeitando as questões éticas da coleta de dados em seres humanos, cujo reuso não estava previsto (R09⁷²).

Identificar quem foi o responsável pela coleta dos dados (R02⁷³), pois esses podem ser coletados por sensores, controles automatizados e instrumentos, além de questionários e entrevistas detalhadas (R04⁷⁴). Se novos dados forem coletados, verificar como integrar aos conjuntos já existentes (R02⁷⁵). Identificar como esses dados foram organizados para que a coleta seja apropriada e os *datasets* possam ser ordenados de maneira a facilitar o reuso (R02⁷⁶).

O pesquisador precisa documentar o processo que usou para coletar dados, contendo a descrição dos próprios dados (quanto, qual formato de dados, qual *software* usar para ler os dados) e as alterações do conjunto de dados (R06⁷⁷). Esse registro viabiliza um relatório histórico de todos os usos e edições dos dados da pesquisa durante um período, garantindo a proveniência dos dados (R06⁷⁸). A forma de apresentação do terceiro requisito é:

⁷⁰ 02-CESSDA\7. Discover\7.3 International surveys - CESSDA TRAINING

⁷¹ 04-DataONE Education Module\02 Data Sharing\L02_Exercise

⁷² 09-MANTRA\9 Data handling tutorials\4 Data handling in NVivo 9\MANTRA-NVIVO-exercise-Unit1

⁷³ 02-CESSDA\1. Plan\1.6.2 DMPQuestionsQuantitativeData

⁷⁴ 04-DataONE Education Module\01 Why Data Management\L01_DataManagement

⁷⁵ 02-CESSDA\20171117DMPQuestionsCESSDAExpertTourGuide

⁷⁶ 02-CESSDA\1. Plan\1.6.2 DMPQuestionsQuantitativeData

⁷⁷ 06-Essentials 4 Data Support\3. Research phase\III.6-Data documentation

⁷⁸ 06-Essentials 4 Data Support\3. Research phase\III.6-Data documentation

CA 3. Aquisição (*Acquisition*)

UR 3.1 Obtenção de Dados (*Obtaining Data*)

UC Ser capaz de detectar quando houver problemas ou não puder ser feito o *download* (total ou parcialmente), realizar pesquisas para obter novos dados. (F02).

UR 3.1 Coleta de Dados (*Collecting Data*)

UC Conhecer os métodos de coleta de dados e as políticas. (F02); Articular um plano para coletar e compartilhar os dados adicionais de suas pesquisas (F06). Executar sistemas para a aquisição e coleta de informações usando a tecnologia da informação adequadamente (por exemplo, processamento de texto, planilhas, sistemas de simulação, bancos de dados). (F04).

- Requisito 4 – **Curadoria** (*Curation*)

O quarto requisito é a **Curadoria** (*Curation*). Os formatos de arquivo impactam na pesquisa e no reúso de dados, eles precisam estar em formatos acessíveis, aceitos pela comunidade (R07⁷⁹), usando formatos amplamente adotados (R07⁸⁰). O pesquisador precisa de familiaridade com diferentes tipos de formatos de dados, podendo ser usados vocabulários controlados para tipos de formatos de dados (R02⁸¹) (R04⁸²) (R08⁸³). Os formatos proprietários têm maior risco de obsolescência, pois dependem de um *software* particular para abertura, leitura e extração de dados (R10⁸⁴).

O pesquisador precisa estar familiarizado com os diversos formatos em relação ao processamento de curto e de longo prazo, em termos de armazenamento e preservação dos dados (R09⁸⁵). Os formatos de curto prazo ou de processamento imediato podem ser o Rich Text Format (* .rtf), MP3, MPEG, JPG, MS Excel (* .xls), Statistical Package for the Social Sciences (SPS) (* .sav, * .por), STATA (* .dta). Os formatos de longo prazo, visando à preservação digital, preferencialmente em formatos abertos são PDF / A, CSV, TIFF, ASCII,

⁷⁹ 07-FOSTER\\foster _ Managing and Sharing Research Data

⁸⁰ 07-FOSTER-The Open Data Handbook\\The Open Data Handbook\\The Open Data Handbook-File Formats

⁸¹ 02-CESSDA\\3. Process\\3.5 File formats and data conversion - CESSDA TRAINING

⁸² 04-DataONE Education Module\\04 Data Entry and Manipulation/

⁸³ 08-Managing and Sharing Research Data handbook\\2 THE RESEARCH DATA LIFECYCLE

⁸⁴ 10-NECDMC\\1 Overview of Research Data Management\\necdmc_module1

⁸⁵ 09-MANTRA\\9 Data handling tutorials\\1 Data handling in SPSS23\\Handling_Data_using_SPSS_workbook_11December2017

Open Document Format (ODF), Extensible Markup Language (XML), XML aberto do Office, JPEG 2000, PNG, SVG, HTML, XHTML, RSS, CSS (R10⁸⁶).

A eventual conversão em si deve ser feita por um pesquisador familiarizado com os dados, pois são necessários padrões para processamento de dados e práticas recomendadas para planilhas e arquivos de dados (R08⁸⁷). A conversão nem sempre é bem-sucedida e pode resultar em perda de dados, o que destaca a importância de armazenar e arquivar dados e documentação brutos (ou seja, não codificados) em formatos não proprietários e adequados à preservação para uso futuro (R09⁸⁸). A fusão de arquivos pode ser realizada, por exemplo, na limpeza, edição de cabeçalhos, exclusão de campos de dados e mescla com outros dados (R06⁸⁹).

O processamento refere-se às transformações que os dados de pesquisa podem sofrer durante os estágios de reuso, como inserção de dados, limpeza de dados; anonimização de informações pessoais quando necessário, descrição, gerenciamento e armazenamento dos dados (R01)⁹⁰. Pode-se incluir a seleção de um subconjunto de dados para análise, mesclando vários conjuntos de dados, manipulação de dados para usabilidade ou transformação de dados (R04⁹¹). Combinação, reprocessamento, (re)agrupamento, conversão, reorganização, adição, fechamento e compactação, interação de dados e migração de dados (R06⁹²). O quarto requisito se apresenta como:

CA 4. Curadoria (*Curation*)

UR 4.1 Conversão (Conversion)

UC Torna-se proficiente na migração de dados de um formato para outro (F01).

Saber converter um formato de dados para outro e ser capaz de estabelecer níveis de acesso a dados. (F02).

UR 4.2 Formatos (Formats)

UC Compreender os riscos e a potencial perda ou corrupção de informações causadas pela alteração dos formatos de dados (F01); Entender quais formatos e tipos de dados são apropriados para diferentes questões de

⁸⁶ 10-NECDMC\6 Data Sharing E Reuse Policies\necdmc_module6

⁸⁷ 08-Managing and Sharing Research Data handbook\5 FORMATTING AND ORGANIZING DATA\CHAPTER 5 UK Data Archive - OUR QUALITY CONTROL

⁸⁸ 9 Data handling tutorials\4 Data handling in NVivo 9\MANTRA-NVIVO-exercise-Unit1

⁸⁹ 06-Essentials 4 Data Support\3. Research phase\III.3-introducing_research_data

⁹⁰ 01-Bodleian\Managing research data and data management planning\Research-data-management-planning-introduction-handbook

⁹¹ 04-DataONE Education Module\09 Analysis and Workflows\L09_AnalysisWorkflows_Handout

⁹² 06-Essentials 4 Data Support\3. Research phase\III.1-Collecting data

pesquisa (F01); Capacidade de reconhecer valores, tipos e formatos de dados de origem (F02).

UR 4.3 Fusão (Merging)

UC Reconhecer e resolver qualquer problema aparente, incompleto ou corrompido dos conjuntos de dados (F01);

UR 4.4 Processamento (Processing)

UC Articular o planejamento e as ações necessárias para realizar a Curadoria de dados (F01).

UR 4.5 Síntese (Synthesize)

UC Sintetizar e representar os resultados da análise de dados de maneira adequada à natureza dos dados, à sua finalidade e ao público alvo da investigação. Escolher métodos adequados de representação de dados (tabelas, gráficos ou similares); sintetizar informações para apoiar argumentos ou gerar novas ideias (F02). Sintetizar criticamente informações novas e complexas de diversas fontes. (F04).

- Requisito 5 – **Gestão de Dados** (*Data Management*)

O quinto requisito é a **Gestão de Dados** (*Data Management*). A primeira ação da gestão de dados é o plano de gestão, em que todas as etapas, tratamento e decisão sobre os dados serão planejadas, como qual padrão de metadados será preciso para a documentação (R01⁹³). Se for necessária alguma medida para proteger informações confidenciais, incluindo a confidencialidade dos participantes do estudo e a proteção da propriedade intelectual (R01⁹⁴, R02⁹⁵).

Os componentes básicos do plano de gestão de dados (DMP) envolvem também a escolha dos tipos de dados, formatos, tamanho do arquivo, padrões e métodos de captura e como os dados devem ser descritos e classificados (por exemplo, quantitativo, qualitativo ou sensível) (R01⁹⁶). Permite planejar o tipo de acesso, o compartilhamento dos dados e a viabilidade de reutilização (R01⁹⁷), a forma de depósito e preservação a longo prazo (R01⁹⁸), bem como as

⁹³ 01-Bodleian\RDM-Kit-session1\DataManagementPlanning-RDMTrainingKit

⁹⁴ 01-Bodleian\RDM-Kit-session1\DataManagementPlanning-RDMTrainingKit

⁹⁵ 02-CESSDA\1. Plan\1.5 European diversity - CESSDA TRAINING

⁹⁶ 01-Bodleian\RDM-Kit-session1\DataManagementPlanning-RDMTrainingKit

⁹⁷ 01-Bodleian\RDM-Kit-session1\DataManagementPlanning-RDMTrainingKit

⁹⁸ 01-Bodleian\RDM-Kit-session1\DataManagementPlanning-RDMTrainingKit

ações de armazenamento de curto prazo e gerenciamento de dados (R01⁹⁹). Os Planos são, em diversos casos, requisitos obrigatórios dos organismos de financiamento (R05¹⁰⁰). O DMP deve ser um documento ativo atualizado durante todo o período do projeto e deve ser tornado público (R02¹⁰¹).

Nem sempre o pesquisador possui todas essas informações no início do projeto. Para tanto, existem ferramentas *online* para criar planos e auxiliá-los ao longo da pesquisa, que inclui o guia *on-line* em como desenvolver um plano de gerenciamento e compartilhamento de dados, uma lista de verificação de itens, exemplos de planos e um resumo dos requisitos dos financiadores, como o *DMP Online Tool*, disponível na Web (R01¹⁰²). A forma de apresentação do quinto requisito é:

CA 5. Gestão de Dados (*Data Management*)

UR 5.1 Plano de Gestão de Dados (*Data Management Plan*)

- UC Identificar os componentes de um plano de gerenciamento de dados (F06).
- Desenvolver planos de gerenciamento de dados (F01).

UR 5.2 Convenção de Nomenclatura (*Naming Conventions*)

- UC Planejar uma estrutura de organização para os dados, usando um sistema de nomeação de arquivos que esteja bem documentado e interoperável (F06).

- Requisito 6 – **Documentação** (*Documentation*)

O sexto requisito é a **Documentação** (*Documentation*): A documentação completa, padronizada e adequada minimiza os riscos dos pesquisadores fazerem má interpretações sobre os dados disponíveis. As respostas dos *datasets* mostraram que a qualidade da documentação é mais relevante que o próprio renome do autor dos dados na decisão de reuso (R11¹⁰³).

Muitos registros são necessários durante um projeto de pesquisa. Esses registros são formalizados na documentação, que no nível do projeto explica os objetivos do estudo, quais são as questões / hipóteses de pesquisa, as metodologias, instrumentos e medidas utilizadas, etc

⁹⁹ 01-Bodleian\\RDM-Kit-session1\\DataManagementPlanning-RDMTrainingKit

¹⁰⁰ 05-DC 101 materials Access and Reuse\\1 DC 101 Access and Use

¹⁰¹ 02-CESSDA\\1. Plan\\1.5 European diversity - CESSDA TRAINING

¹⁰² 01-Bodleian\\Managing research data and data management planning\\Research-data-management-planning-introduction-handbook

¹⁰³ 10-NECDMC\\2 Types, Formats, and Stages of Data\\necdmc_module2

(R02¹⁰⁴). A documentação no nível de coleta de dados da pesquisa também contém informações detalhadas sobre a coleta, os respondentes, as ferramentas de medição, a manipulação de dados etc. Inclui guias do usuário, questionários de pesquisa, agendas de entrevistas e notas de trabalho de campo (R02¹⁰⁵).

O pesquisador precisa identificar a documentação criada durante as fases do projeto e como essa documentação foi estruturada. São informações necessárias para interpretar e ler os dados corretamente para o reuso (R11¹⁰⁶). Também inclui acordos de licença e todos os materiais obtidos na negociação original e no depósito de dados, bem como informações pós-depósito criadas durante as atividades de preservação e atualização. Os termos dos metadados e da documentação geralmente são usados de forma interoperável (R10¹⁰⁷).

O Pesquisador deve procurar documentação de dados de alta qualidade em repositórios confiáveis, preferencialmente pelos dados que sejam acompanhados por documentação em nível de projeto e em nível de coleta de dados (R02¹⁰⁸). Essa documentação geralmente pode ser encontrada em documentos chamados *Guia do Usuário*, notas de trabalho de campo, relatório técnico ou arquivos *Leia-me* (R02¹⁰⁹). Para documentação no nível de dados, a incorporação dessa documentação dentro dos arquivos de dados é uma boa prática recomendada (R10¹¹⁰).

A documentação traz a descrição dos próprios dados (quanto, qual formato de dados, qual *software* usar para ler os dados) e das alterações do conjunto de dados (R06¹¹¹). Alguns exemplos de documentação de dados são: cadernos de laboratório, questionários de protocolos experimentais, livros de códigos e dicionários de dados nas Ciências Sociais, informações sobre fontes de dados derivados ou digitalizados (R10¹¹²). Isso é usado para criar um relatório histórico de todos os usos e edições dos dados da pesquisa durante um período de tempo e esse é registro de proveniência dos dados (R06¹¹³). A documentação adequada garante que os dados da pesquisa sejam rastreáveis e sejam inequivocamente entendidos e usados pelos usuários

¹⁰⁴ 02-CESSDA\7. Discover\7.1 The process of data discovery - CESSDA TRAINING

¹⁰⁵ 02-CESSDA\7. Discover\7.1 The process of data discovery - CESSDA TRAINING

¹⁰⁶ 10-NECDMC\2 Types, Formats, and Stages of Data\necdmc_module2

¹⁰⁷ 10-NECDMC\2 Types, Formats, and Stages of Data\necdmc_module2

¹⁰⁸ 02-CESSDA\7. Discover\7.1 The process of data discovery - CESSDA TRAINING

¹⁰⁹ 02-CESSDA\7. Discover\7.1 The process of data discovery - CESSDA TRAINING

¹¹⁰ 10-NECDMC\2 Types, Formats, and Stages of Data\necdmc_module2

¹¹¹ 06-Essentials 4 Data Support\3. Research phase\III.6-Data documentation

¹¹² 10-NECDMC\2 Types, Formats, and Stages of Data\necdmc_module2

¹¹³ 06-Essentials 4 Data Support\3. Research phase\III.6-Data documentation

atuais e futuros (incluindo o pesquisador), além de garantir a preservação bem-sucedida a longo prazo das descobertas da pesquisa (R06¹¹⁴).

As práticas de documentação de dados variam de acordo com a disciplina, e embora práticas específicas difiram bastante, um aspecto é universal: nomeação de arquivos e convenções de organização (R10¹¹⁵). Arquivos bem organizados e com nomes claros podem ajudar a garantir a recuperação mais eficiente (R10¹¹⁶). O padrão de documentação estabelecido pela Data Documentation Initiative (DDI) é um guia que vincula a documentação às principais fases do ciclo de vida dos dados de pesquisa, e inclui: Descoberta e Planejamento, Coleta inicial de dados, Preparação e Análise de dados, Publicação e Compartilhamento e Gerenciamento de longo prazo, e pode ser usado para compreender a documentação dos dados a serem reutilizados (R11¹¹⁷). Como forma de representação tem-se:

CA 6. Documentação (*Documentation*)

UR 6.1 Descrição de Conjunto de Dados (*Data Sets Description*)

UC Organizar os dados coletados, rastrear versões, compreender os metadados e a documentação dos dados (F06). Nomear as partes interessadas de dados (F06).

Ajudar outros pesquisadores a entender os dados. (F06)

- Requisito 7 – **Licenças** (*Licences*)

O sétimo requisito trata das **Licenças** (*Licences*). Esse requisito é um dos mais abordados nos recursos de aprendizado e trata, neste contexto, das competências de citação, de Propriedade Intelectual e de uso ético dos dados (R01¹¹⁸) (R03¹¹⁹) (R05¹²⁰) (R08¹²¹). O letramento em dados passa, obrigatoriamente, por esse requisito. A adequada atribuição de créditos aos pesquisadores que geraram os dados é um fator relevante para a cultura do compartilhamento (R02¹²²) (R09¹²³). No meio acadêmico a recompensa está baseada em

¹¹⁴ 06-Essentials 4 Data Support\3. Research phase\III.6-Data documentation

¹¹⁵ 10-NECDMC\2 Types, Formats, and Stages of Data\necdmc_module2

¹¹⁶ 10-NECDMC\2 Types, Formats, and Stages of Data\necdmc_module2

¹¹⁷ 11-Research Data Management and Sharing - Coursera\1-Understanding Research Data\1-What Are Data

¹¹⁸ 01-Bodleian\Kit-session4-Ethics_Copyright\Ethics_Copyright-RDMTrainingKit

¹¹⁹ 03-Data Management Course\1. Introduction to Data Management\1.2 Syllabus - Data Management Course

¹²⁰ 05-DC 101 materials Access and Reuse\1 DC 101 Access and Use

¹²¹ 08-Managing and Sharing Research Data handbook\2 THE RESEARCH DATA LIFECYCLE\managingsharing

¹²² 02-CESSDA\6. Archive & Publish\6.5 Citing your data - CESSDA TRAINING

¹²³ 09-MANTRA\5 Documentation, metadata, citation\Documentation, metadata, citation20

reconhecimento e reputação, portanto, os pesquisadores precisam ser reconhecidos também pelos seus dados publicados com vistas à reutilização, concedendo o crédito acadêmico e a atribuição normativa e legal a todos os contribuintes dos dados, assim como já é uma prática adotada para os documentos bibliográficos (R02¹²⁴) (R08¹²⁵).

A citação de dados deve incluir, principalmente, um método persistente - identificador persistente como o Digital Object Identifier (DOI) (R04¹²⁶) (R05¹²⁷) (R08¹²⁸). Um identificador persistente é uma identificação exclusiva que permite aos conjuntos de dados serem encontrados e citados (R05¹²⁹) (R08¹³⁰). O pesquisador que pretende reusar dados precisa reconhecer localizadores de dados persistentes comuns, como DOI, Handle, chave de recurso de arquivamento (ARK), identificador universal (UUID) (R08¹³¹).

Há padrões de citação de dados, mas existem elementos de citação mínima de dados (R02¹³²) como autor, data de publicação, título, editor ou local, o identificador persistente (DOI), podendo também conter a localização (endereço da *web*), número da versão ou edição, a versão do conjunto de dados usado no estudo, a data de acesso e o nome do recurso ou subconjunto de dados utilizados (R03¹³³). A descrição é baseada na organização do conjunto de dados maior, local de publicação, cidade e estado e país da distribuidor dos dados (R04¹³⁴).

O pesquisador que pretende reusar dados deve reconhecer esses elementos de metadados, úteis para citação de dados. Inclusive, há alguns aplicativos que suportam a criação de metadados para conjuntos de dados (R05¹³⁵). Existem tecnologias disponíveis para apoiar a citação de dados, como o DataCite DOIs: DataCite Consortium (fornece interfaces humanas e de máquina para usuários finais simples na administração de registros DOI) (R05¹³⁶). O DataCite também coleta metadados sobre cada conjunto de dados que registra, que podem ser, posteriormente, pesquisados pela interface *Web* ou colhidos usando protocolo interoperável

¹²⁴ 02-CESSDA\6. Archive & Publish\6.5 Citing your data - CESSDA TRAINING

¹²⁵ 04-DataONE Education Module\08 Data Citation\L08_DataCitation

¹²⁶ 04-DataONE Education Module\08 Data Citation\L08_DataCitation

¹²⁷ 05-DC 101 materials Access and Reuse\3 How_to_Cite_Link

¹²⁸ 04-DataONE Education Module\08 Data Citation\L08_DataCitation

¹²⁹ 05-DC 101 materials Access and Reuse\1 DC 101 Access and Use

¹³⁰ 04-DataONE Education Module\08 Data Citation\L08_DataCitation

¹³¹ 04-DataONE Education Module\08 Data Citation\L08_DataCitation

¹³² 02-CESSDA\7. Discover\7.6 Access, use and cite data - CESSDA TRAINING

¹³³ 03-Data Management Course\5. Data Sharing and Re-use\Flipped Data Management- session 4

¹³⁴ 04-DataONE Education Module\08 Data Citation\L08_DataCitation

¹³⁵ 05-DC 101 materials Access and Reuse\3 How_to_Cite_Link

¹³⁶ 05-DC 101 materials Access and Reuse\3 How_to_Cite_Link

(R05¹³⁷) (R05¹³⁸). Também se deve fazer o uso de palavras-chave padronizadas para descrever seus dados, disponíveis em tesouros e vocabulários controlados (R05¹³⁹).

Os pesquisadores devem, ao reusar dados, compreender como as leis de propriedade de propriedade intelectual e / ou de proteções de direitos autorais se relacionam com dados de pesquisa que pretendem utilizar, identificar quem possui os direitos autorais ou de propriedade intelectual dos dados (R02¹⁴⁰) (R07¹⁴¹) (R08¹⁴²) (R09¹⁴³).

O pesquisador precisa compreender como lidar com questões éticas ou de privacidade (R03¹⁴⁴) (como a anonimização de dados, por exemplo – R05¹⁴⁵; R09¹⁴⁶). É necessário explicar e avaliar possíveis questões legais relacionadas ao reuso de dados, das reivindicações de direitos autorais, das licenças necessárias para uso, e inclusive de eventuais cobranças monetárias por dados (R05¹⁴⁷). Na fase de planejamento da pesquisa que envolva o reuso de dados, é preciso avaliar todas as questões legais, não só para o conjunto de dados, mas para o reuso de códigos de *softwares* (R05¹⁴⁸).

As licenças e suas permissões devem ser conhecidas e compreendidas, e ao utilizar um conjunto de dados, reconhecer qual licença foi atribuída pelo autor dos dados (R05¹⁴⁹). As licenças de reutilização podem ser, por exemplo, uma das licenças disponíveis no Creative Commons ou Open Data Commons (R09¹⁵⁰), por isso é necessário que o pesquisador esteja familiarizado com elas (R05¹⁵¹).

O pesquisador deve reconhecer quem tem a propriedade primária dos dados e quais direitos devem ser considerados quando decidir sobre o reuso, o processamento, o gerenciamento e a re-disseminação dos dados (R09¹⁵²). A propriedade e os direitos, normalmente expressos nas licenças atribuídas pelos autores dos dados é que determinarão

¹³⁷ 05-DC 101 materials Access and Reuse\\8 data_management_frameworks

¹³⁸ 05-DC 101 materials Access and Reuse\\3 How_to_Cite_Link

¹³⁹ 05-DC 101 materials Access and Reuse\\3 How_to_Cite_Link

¹⁴⁰ 02-CESSDA\\7. Discover\\7.6 Access, use and cite data - CESSDA TRAINING

¹⁴¹ 07-FOSTER\\foster-Cite data _ Find, reuse and cite data _ Library _ University of Leeds

¹⁴² 08-Managing and Sharing Research Data handbook\\11 PUBLISHING AND CITING RESEARCH DATA\\CHAPTER 11 publishingcitingdata

¹⁴³ 09-MANTRA\\7 Data protection, rights & access\\Data protection, rights and access20

¹⁴⁴ 03-Data Management Course\\4. Data Access and Ownership\\4. Data Access and Ownership - Data Management Course

¹⁴⁵ 05-DC 101 materials Access and Reuse\\8 data_management_frameworks

¹⁴⁶ 09-MANTRA\\9 Data handling tutorials\\4 Data handling in NVivo 9\\MANTRA-NVIVO-exercise-Intro

¹⁴⁷ 05-DC 101 materials Access and Reuse\\1 DC 101 Access and Use

¹⁴⁸ 05-DC 101 materials Access and Reuse\\8 data_management_frameworks

¹⁴⁹ 05-DC 101 materials Access and Reuse\\7 IPR in Databases

¹⁵⁰ 09-MANTRA\\2 Data management plans\\Data management plans17d

¹⁵¹ 05-DC 101 materials Access and Reuse\\8 data_management_frameworks

¹⁵² 09-MANTRA\\7 Data protection, rights & access\\Data protection, rights and access20

como os dados podem ser gerenciados no futuro, portanto, essas informações devem ser documentadas no início de um projeto, incluídas no plano de gestão de dados (R09¹⁵³). A forma de representação é:

CA 7. Licenças (*Licences*)

UR 7.1 Citação (*Citation*)

UC Saber como citar fontes de dados (F02). Produzir citações precisas para os dados (F02). Usar identificadores persistentes para citar dados da pesquisa (F03). Citar seus dados em um formato adequadamente estruturado, de acordo com os padrões emergentes (F06).

UR 7.2 Propriedade Intelectual (*Intellectual Property Rights*)

UC Entender as implicações éticas e legais do uso de dados em termos de propriedade intelectual, privacidade e citações (F02). Entender as possíveis preocupações de propriedade intelectual ao liberar seus dados para um público mais amplo. (F06). Compreender os direitos autorais no conteúdo de Dados abertos, Acesso aberto, Ciência aberta (F04). Identificar as leis locais de direitos autorais (F03). Interpretar licenças de conteúdo e direitos autorais (F03). Entender as leis de direitos autorais relevantes para sua disciplina (F03).

UR 7.3 Uso Ético dos Dados (*Using Data Ethically*)

UC Reconhecer a fonte quando obtida ou formulada por terceiros; registrar ou comunicar os resultados de maneira ética (F02). Ter consciência de questões relacionadas aos direitos de outros pesquisadores, sujeitos da pesquisa e de outras pessoas que possam ser afetadas pela pesquisa (F04). Compreender e aplicar os códigos de conduta e diretrizes relevantes para a condução ética da pesquisa; procurar aconselhamento do orientador. (F04)

- Requisito 8 – **Metadados** (*Metadata*)

O oitavo requisito são os **Metadados** (*Metadata*). Metadados são as informações que descrevem aspectos significativos, por exemplo, conteúdo, contexto e estrutura das informações

¹⁵³ 09-MANTRA\\7 Data protection, rights & access\\Data protection, rights and access20

de um registro (R08¹⁵⁴). Eles permitem a descoberta de registros, o gerenciando, o acesso e a preservação. Os metadados podem existir em vários níveis, desde a coleta de dados até as variáveis individuais de cada arquivo na coleção (R07¹⁵⁵).

Metadados compõem a documentação sobre os dados, são elementos estruturados que descrevem os dados da pesquisa. A documentação e os metadados permitem que os dados sejam compreendidos e descobertos para o reuso (R07¹⁵⁶). Os metadados do processo registram informações sobre o processo (análise, organização de dados, gráficos) usados para obter saídas de dados (R07¹⁵⁷). Já os metadados relacionados a proveniência dos dados registram as origens dos dados, acompanham os dados durante todo o ciclo de vida e permitem a replicação e reprodutibilidade. Os metadados de preservação, como os metadados de fixidez, de preservação e de gerenciamento de registros (R08¹⁵⁸) são utilizados com vistas ao armazenamento (R08¹⁵⁹).

Os metadados devem acompanhar os dados para que sejam interpretados corretamente, pois contém detalhes contextuais críticos que permite entender e usar os dados, incluindo detalhes espaciais e temporais, instrumentos, parâmetros, unidades e arquivos (R04¹⁶⁰), ou seja, informações críticas sobre o conteúdo de um conjunto de dados (R04¹⁶¹). Também descrevem a qualidade, a condição e as outras características de um conjunto de dados, permitem que os dados sejam descobertos, acessados e reutilizados (R04¹⁶²).

As citações de dados facilitam o acesso aos dados em si e aos metadados, a documentação e outros materiais associados, conforme necessário para pessoas e máquinas para usar os dados referenciados (R05¹⁶³). Dados pessoais sensíveis ou dados comercialmente sensíveis podem não ser viáveis de reuso, pois não estariam compartilhados (R05¹⁶⁴). Os padrões de metadados geralmente têm uma estrutura hierárquica (R08¹⁶⁵) e projetada para ser

¹⁵⁴ 08-Managing and Sharing Research Data handbook\6 STORING AND TRANSFERRING
DATA\CHAPTER 6 cd062-preservationpolicy

¹⁵⁵ 07-FOSTER-The Open Data Handbook\07-Full\fooster _ Managing and Sharing Research Data

¹⁵⁶ 07-FOSTER-The Open Data Handbook\07-Full\fooster _ Managing and Sharing Research Data

¹⁵⁷ 07-FOSTER-The Open Data Handbook\07-Full\fooster _ Managing and Sharing Research Data

¹⁵⁸ 08-Managing and Sharing Research Data handbook\6 STORING AND TRANSFERRING
DATA\CHAPTER 6 cd062-preservationpolicy

¹⁵⁹ 08-Managing and Sharing Research Data handbook\6 STORING AND TRANSFERRING
DATA\CHAPTER 6 cd062-preservationpolicy

¹⁶⁰ 04-DataONE Education Module\03 Data Management Planning\L03_DataManagement_Handout

¹⁶¹ 04-DataONE Education Module\07 Metadata\L07_Metadata

¹⁶² 04-DataONE Education Module\L07_DefiningMetadata_Handout

¹⁶³ 05-DC 101 materials Access and Reuse\3 How_to_Cite_Link

¹⁶⁴ 07-FOSTER-The Open Data Handbook\07-Full\fooster _ Managing and Sharing Research Data

¹⁶⁵ 08-Managing and Sharing Research Data handbook\6 STORING AND TRANSFERRING
DATA\CHAPTER 6 cd062-preservationpolicy

legível por máquina. Pode ser usada, por exemplo, para criar um catálogo de conjunto de dados para um repositório (R08¹⁶⁶).

As informações relevantes variam entre as disciplinas, portanto, muitas áreas acadêmicas têm padrões de metadados ou formatos recomendados nos quais os dados sobre conjuntos de dados devem ser apresentados. O pesquisador deve conhecer essas informações com antecedência, para garantir que esteja compreendendo as informações relevantes (R1¹⁶⁷). Os metadados são um subconjunto (que implica processamento por computador, embora também possa ser legível por humanos (R01¹⁶⁸) da documentação (para ser lido por humanos (R01¹⁶⁹), descrevendo e contextualizando os dados em detalhes(R07¹⁷⁰). Existem vários padrões de metadados, desenvolvidos e assinados por certas comunidades (R06¹⁷¹), como por exemplo o Data Documentation Initiative¹⁷² (DDI¹⁷³) (R08¹⁷⁴), o Dublin Core Metadata Element Set¹⁷⁵ (R03¹⁷⁶), o Dublin Core Metadata Initiative (R06¹⁷⁷), o Research Description Framework (RDF) (R08¹⁷⁸), o Metadata Encoding and Transmission Standard (METS) (R08¹⁷⁹) e o Preservation Metadata Maintenance Activity (PREMIS) (R08¹⁸⁰). Como representação tem-se:

CA 8. Metadados (*Metadata*)

UR 8.1 Descrição de Metadados (*Metadata Description*)

UC Saber o que são metadados (F02). Estar ciente da necessidade de salvar os dados selecionados ou gerados e de dados descritivos ou outros associados, para a devida identificação, gerenciamento e subsequente reutilização. (F02).

¹⁶⁶ 08-Managing and Sharing Research Data handbook\4 DOCUMENTING AND PROVIDING CONTEXT FOR DATA\CHAPTER 4 documentingdata

¹⁶⁷ 01-Bodleian\Managing research data and data management planning\Research-data-management-planning-introduction-handbook

¹⁶⁸ 01-Bodleian\Kit-session2-Organising_DocumentingData\Organising_DocumentingData-RDMTrainingKit

¹⁶⁹ 01-Bodleian\Kit-session2-Organising_DocumentingData\Organising_DocumentingData-RDMTrainingKit

¹⁷⁰ 07-FOSTER-The Open Data Handbook\07-Full\fooster_Managing and Sharing Research Data

¹⁷¹ 06-Essentials 4 Data Support\4. User phase\IV.4-Metadata

¹⁷² <http://www.ddialliance.org>

¹⁷³ <http://www.ddialliance.org/>

¹⁷⁴ 08-Managing and Sharing Research Data handbook\6 STORING AND TRANSFERRING DATA\CHAPTER 6 cd062-preservationpolicy

¹⁷⁵ <http://dublincore.org/documents/dces>

¹⁷⁶ 06-Essentials 4 Data Support\4. User phase\IV.4-Metadata

¹⁷⁷ 06-Essentials 4 Data Support\4. User phase\IV.4-Metadata

¹⁷⁸ 08 Sharing, preservation & licensing\Sharing, preservation, and licensing19

¹⁷⁹ 08-Managing and Sharing Research Data handbook\4 DOCUMENTING AND PROVIDING CONTEXT FOR DATA\CHAPTER 4 documentingdata

¹⁸⁰ 08-Managing and Sharing Research Data handbook\4 DOCUMENTING AND PROVIDING CONTEXT FOR DATA\CHAPTER 4 documentingdata

Utilizar os metadados para facilitar o entendimento de possíveis problemas com conjuntos de dados. (F01).

UR 8.2 Padrões de Metadados (Metadata Standards)

UC Saber como identificar padrões e práticas aplicáveis de metadados. (F02).

Conhecer a finalidade dos metadados. (F04).

- **Requisito 9 – Ciência Aberta** (*Open Science*)

O nono requisito é a **Ciência Aberta** (*Open Science*). Um princípio fundamental da ciência aberta é garantir a disponibilidade pública e a reutilização de dados científicos (R10¹⁸¹) e esses recursos ajudam o pesquisador a incorporar a Ciência Aberta nas práticas de pesquisa de trabalho (R07¹⁸²), promovendo maior transparência dos métodos de pesquisa e de avaliação (R07¹⁸³).

É importante para o pesquisador compreender as controvérsias em torno da Ciência Aberta (R05¹⁸⁴), que mesmo que esta seja parte das boas práticas científicas (R07¹⁸⁵), levando às pesquisas mais eficientes e às inovações (R10¹⁸⁶), existe uma preocupação com o potencial de indivíduos, instituições ou nações inteiras lucrarem com o trabalho de outras pessoas. Também muitos pesquisadores temem que seu trabalho possa ser mau interpretado ou editado e isso pode diminuir suas contribuições para a Ciência, em vez de aprimorá-la (R10¹⁸⁷). Porém, dados acessados mais facilmente são mais prováveis de reuso, aumentando o impacto no trabalho do autor dos dados (R02¹⁸⁸).

O pesquisador deve considerar a possibilidade de publicação em periódicos de dados revisados por pares, projetados para documentar e publicar de maneira abrangente os conjuntos de dados depositados e facilitar sua exploração on-line (R02¹⁸⁹). O pesquisador pode usar periódicos de dados especializados que publicam descrições de conjuntos de dados

¹⁸¹ 10-NECDMC\1 Overview of Research Data Management\necdmc_module1

¹⁸² 07-FOSTER-The Open Data Handbook\07-Full\fooster_Best Practice in Open Research

¹⁸³ 07-FOSTER-The Open Data Handbook\07-Full\Future Proof and FAIR Research Data

¹⁸⁴ 05-DC 101 materials Access and Reuse\8 data_management_frameworks

¹⁸⁵ 07-FOSTER-The Open Data Handbook\07-Full\Future Proof and FAIR Research Data

¹⁸⁶ 10-NECDMC\6 Data Sharing E Reuse Policies\necdmc_module6

¹⁸⁷ 10-NECDMC\6 Data Sharing E Reuse Policies\necdmc_module6

¹⁸⁸ 02-CESSDA\6. Archive & Publish\6.8 Promoting your data - CESSDA TRAINING

¹⁸⁹ 02-CESSDA\6. Archive & Publish\6.8 Promoting your data - CESSDA TRAINING

cientificamente valiosos e textos de pesquisa sobre o compartilhamento e o reúso de dados (R02¹⁹⁰).

Quanto ao uso e o reúso de dados abertos, um conjunto de princípios foi estabelecido para garantir a usabilidade dos dados no longo prazo, são eles: os dados precisam ser localizáveis, acessíveis, interoperáveis e reutilizáveis (*Findable, Accessible, Interoperable and Reusable* – FAIR) (R07¹⁹¹).

Os princípios FAIR recomendam o uso de padrões de dados compartilhados para a representação de dados e o uso de vocabulários e ontologias para representar os valores. Vocabulários e ontologias são frequentemente definidos nas comunidades de pesquisa e é uma maneira inequívoca de adicionar significado semântico aos seus dados (R07¹⁹²). Os princípios orientadores da FAIR consistem em 15 facetas que descrevem um *continuum* de reutilização crescente (R02¹⁹³). Os dados não devem ser apenas FAIR para humanos, mas também para máquinas, permitindo, por exemplo, pesquisa e acesso automatizados aos dados (R02¹⁹⁴). FAIR não significa que os dados precisam ser abertos, deve haver metadados mesmo para dados não acessíveis em acesso aberto (R02¹⁹⁵). A representação para a Ciência Aberta é:

CA 9. Ciência Aberta (*Open Science*)

UR 9.1 Conceito de Ciência Aberta (*Concept of Open Science*)

UC Entender significado de ciência aberta e dados abertos (F02). Entender o valor do acesso aberto dos resultados da pesquisa aos pesquisadores e à sociedade em geral (F04).

UR 9.2 Acesso Aberto (*Open Access*)

UC Identificar ferramentas e infraestrutura eletrônica para Acesso Aberto (F03).

UR 9.3 Periódicos de Dados Abertos (*Open Data Journals*)

UC Identificar os periódicos abertos de dados existentes (F03). Preparar uma publicação para um Open Data Journal (F03).

UR 9.3 Uso E Reúso de Dados Abertos (*Open Data Use and Reuse*)

¹⁹⁰ 02-CESSDA\7. Discover\7.1 The process of data discovery - CESSDA TRAINING

¹⁹¹ 07-FOSTER-The Open Data Handbook\07-Full\fooster_Best Practice in Open Research

¹⁹² 07-FOSTER-The Open Data Handbook\07-Full\A FAIRy tale

¹⁹³ 02-CESSDA\1. Plan\1.4 FAIR data - CESSDA TRAINING

¹⁹⁴ 02-CESSDA\1. Plan\1.4 FAIR data - CESSDA TRAINING

¹⁹⁵ 02-CESSDA\1. Plan\1.4.1 20170707_How_FAIR_are_your_data_Jones

UC Identificar as ferramentas do *open source* (códigos abertos) para cada etapa do ciclo de vida da pesquisa (F03); Definir a relevância das ferramentas do *open source* para Reprodutibilidade e Integridade da Pesquisa (F03).

- Requisito 10 – **Política** (*Policy*)

O décimo requisito é a **Política** (*Policy*). O acesso, o uso e a reutilização de dados devem ser realizados com total reconhecimento dos requisitos contidos nas políticas (R05¹⁹⁶). Por exemplo, podem haver restrições legais que significam que os dados não podem ser compartilhados, casos em que o proprietário dos dados pode não permitir a reutilização dos mesmos ou direitos de propriedade intelectual para alguns materiais serem tão restritivos que o acesso à esse material fica inviabilizado (R01¹⁹⁷) (R05¹⁹⁸). Os requisitos legais podem ser dos órgãos de financiamento da pesquisa, ou de legislações específicas sobre confidencialidade e privacidade, e direitos de propriedade intelectual e de gerenciamento de ativos digitais, como a legislação relacionada à confidencialidade e à privacidade, como atos de proteção de dados e atos de liberdade de informação que podem restringir o acesso, o uso e a reutilização (R10¹⁹⁹).

Porém, nota-se que nas políticas institucionais de agências de fomento, o acesso aos dados, para que possam ser usados por outros que não os seus criadores e reutilizados no futuro, é cada vez mais exigido por esses organismos (R01²⁰⁰). Como as políticas de propriedade intelectual variam de acordo com a instituição, familiarize-se com as políticas de sua instituição e do órgão financiador de pesquisa quanto ao reuso de dados, buscando identificar quais licenças podem ser usadas (R10²⁰¹). A representação proposta:

CA 10. Políticas (*Policy*)

UR 10.1 Políticas de Financiamento (*Funding Policies*)

UC Ter ciência do papel dos dados na sociedade, como eles são gerados e por quem, e seus possíveis aplicações, bem como as implicações de seu uso. (F02). Ter amplo conhecimento das principais fontes de

¹⁹⁶ 05-DC 101 materials Access and Reuse\\1 DC 101 Access and Use

¹⁹⁷ 01-Bodleian\\Introducing MANTRA for Research Data Management\\Introducing_MANTRA_for_research_data_management_2018

¹⁹⁸ 05-DC 101 materials Access and Reuse\\1 DC 101 Access and Use

¹⁹⁹ 10-NECDMC\\5 Legal and Ethical Considerations for Research Data\\necdmc_lessonplan5

²⁰⁰ 01-Bodleian\\Introducing MANTRA for Research Data

Management\\Introducing_MANTRA_for_research_data_management_2018

²⁰¹ 10-NECDMC\\5 Legal and Ethical Considerations for Research Data\\necdmc_lessonplan5

financiamento relevantes e procedimentos de solicitação de subsídios. (F04).

- Requisito 11 – **Qualidade** (*Quality*)

O décimo primeiro requisito é a **Qualidade** (*Quality*): Há as preocupações dos pesquisadores sobre qualidade e confiabilidade dos dados a serem reutilizados (R04²⁰²). Dados disponíveis na própria organização podem influenciar positivamente as intenções de reutilização de dados. Mecanismos de controle de qualidade de repositórios, como os serviços de verificação da qualidade dos repositórios e da proveniência dos dados também são capazes de avaliar se os dados atendem aos pesquisadores para o reuso (R09²⁰³).

Garantia e controle de qualidade são estratégias para evitar erros ao inserir um conjunto de dados, monitorando e mantendo a qualidade dos dados ao longo do projeto (R04)²⁰⁴, em todas as etapas do ciclo de pesquisa (R04²⁰⁵). As práticas recomendadas para garantia de qualidade são o controle de versão e autenticidade e o gerenciamento de versão e edição (R02²⁰⁶), também quanto à integridade, compreensibilidade e validação dos arquivos de dados (R02²⁰⁷).

A garantia de qualidade usa estratégias de validação para verificar o volume de dados, formatos de arquivo, número de arquivos, organização de dados - estrutura de diretórios, convenções de nomenclatura de arquivos, estrutura de arquivos, conteúdo dos dados - nomes e descrições de variáveis, esquemas de classificação utilizados, métodos de análise de dados; algoritmos usados para processar dados, Confidencialidade, condições de acesso e uso (R05²⁰⁸), garantindo que os dados sejam limpos e verificados (R08²⁰⁹).

Mesmo que medidas de garantia de qualidade tenham sido tomadas antes do compartilhamento de dados externo (R10²¹⁰), o pesquisador também pode verificar se os dados

²⁰² 04-DataONE Education Module\05 Data Quality Control and Assurance\L05_DataQualityControlAssurance

²⁰³ 09-MANTRA\2 Data management plans\Data management plans11

²⁰⁴ 04-DataONE Education Module\05 Data Quality Control and Assurance\L05_DataQualityControlAssurance

²⁰⁵ 04-DataONE Education Module\05 Data Quality Control and Assurance\L05_DataQualityControlAssurance

²⁰⁶ 02-CESSDA\3. Process\3.6 Data authenticity - CESSDA TRAINING

²⁰⁷ 02-CESSDA\6. Archive & Publish\6.4 Publishing with CESSDA archives - CESSDA TRAINING

²⁰⁸ 05-DC 101 materials Access and Reuse\8 data_management_frameworks

²⁰⁹ 08-Managing and Sharing Research Data handbook\5 FORMATTING AND ORGANIZING DATA\CHAPTER 5 formattingorganising

²¹⁰ 10-NECDMC\Extras\necdmc_researchcase_combining

foram revisados por pares (R08²¹¹). Ele também pode procurar conhecer os protocolos de garantia de qualidade (R11²¹²), como o Quality Assurance Framework of the European Statistical System (Quadro de Garantia da Qualidade do Sistema Estatístico Europeu (R02²¹³)). Como representação propõe-se:

CA 11. Qualidade (*Quality*)

UR 11.1 Garantia da Qualidade (*Quality Assurance*)

UC Avaliar a qualidade dos dados com base na confiabilidade e autoridade. (F02).
Avalia a qualidade, a integridade e a autenticidade das informações ou dos dados da pesquisa primária e secundária. (F04).

- Requisito 12 – **Reúso** (*Reuse*)

O décimo segundo requisito é o **Reúso** (*Reuse*): Ainda que o reúso de dados tenha particularidades disciplinares e motivacionais, a efetividade e a boa relação custo-benefício deve ser considerada. Algumas boas práticas em documentação de dados, metadados e citação podem contribuir para alcançar um melhor padrão de reprodutibilidade da pesquisa, um princípio fundamental do método científico (R09²¹⁴).

Outro aspecto de apoio à reutilização é a garantia do consentimento informado aos participantes sobre o compartilhamento dos dados, desde que a pesquisa envolva seres humanos (R07²¹⁵). A legislação de proteção deve ser atendida para permitir o reúso, portanto deve ser garantido que os entrevistados sejam protegidos, por meio de segurança técnica, administrativa e física, com todos os dados pessoais anonimizados (R02²¹⁶).

Foi desenvolvido um padrão para a pesquisa reprodutível (Reproducible Research Standard - RRS) em resposta às situações em que a maioria dos componentes necessários para a reprodução dos resultados da pesquisa (como código, parâmetros usados, conjunto de dados,

²¹¹ 08-Managing and Sharing Research Data handbook\8 RIGHTS RELATING TO RESEARCH DATA\CHAPTER 8 UK Data Service 3

²¹² 11-Research Data Management and Sharing - Coursera\1-Understanding Research Data\1-Data Management Across the Research Lifecycle

²¹³ 02-CESSDA\3. Process\3.7 Wrap up_ Data quality - CESSDA TRAINING

²¹⁴ 09-MANTRA\5 Documentation, metadata, citation\Documentation, metadata, citation2

²¹⁵ 07-FOSTER-The Open Data Handbook\07-Full\fooster _ Managing and Sharing Research Data

²¹⁶ 02-CESSDA\2. Organise & Document\1.6.2 DMPQuestionsQuantitativeData

detalhes da aquisição, documentação e qualquer conhecimento utilizado no experimento) geralmente permaneciam não publicados (R05²¹⁷).

Este padrão segue os princípios da integridade científica e incentiva a pesquisa reprodutível, pois se configura como uma abordagem sistemática em vista da reutilização (R06²¹⁸). A integridade científica é um termo genérico para a conduta que um pesquisador deve cumprir para assegurar que um estudo possa ser validado e reproduzido. Vários movimentos surgem dentro da própria comunidade de pesquisa que defendem a pesquisa reprodutível (R06²¹⁹). Entretanto, a viabilidade da reprodutibilidade deve ser considerada ainda na fase de planejamento, pois certas opções que levam permitem-na precisam estar arraigadas em seu projeto de pesquisa e no plano de gerenciamento de dados. (R06²²⁰), documentando bem o processo e todos os recursos necessários para reproduzir a pesquisa (dados, código e *software*) deve estar acessível ao público (R06²²¹). Os repositórios de dados já estão expandindo seu papel para garantir a qualidade da pesquisa reprodutível (R09²²²). À medida que a Ciência se torna cada vez mais baseada na computação, e não na experimentação, uma lacuna de credibilidade está surgindo da falta de acesso aos dados e códigos necessários para reproduzir os resultados. Portanto, fontes complementares aos artigos publicados, que incluem todos os dados, documentação e códigos, são necessários para reproduzir os resultados (R09²²³).

O pesquisador precisa estar em conformidade com diferentes requisitos legais dos Estados Membros da União Europeia, por exemplo, e consciente do impacto do Regulamento Geral de Proteção de Dados (General Data Protection Regulation - GDPR²²⁴) no gerenciamento de dados de pesquisa. Há regulações em outros países, como nos Estados Unidos, a Califórnia Consumer Protection Act (CCPA) e no Brasil, a Lei Geral de Proteção de Dados (LGDP), que passarão a vigorar em 2020. Todas essas legislações impõem condições para a coleta e o processamento (o que inclui o simples armazenamento) de dados pessoais, como a obtenção de consentimento informado, a anonimização e a regulação do acesso aos dados (R02²²⁵). Informações passíveis de serem identificadas estão excluídas do compartilhamento, entretanto, até dados sensíveis podem ser compartilhados se forem adotados procedimentos e precauções

²¹⁷ 05-DC 101 materials Access and Reuse\6 Creative Commons

²¹⁸ 06-Essentials 4 Data Support\2. Planning phase\II - Planning phase

²¹⁹ 06-Essentials 4 Data Support\2. Planning phase\II.1-Scientific integrityA

²²⁰ 06-Essentials 4 Data Support\2. Planning phase\II.1-Scientific integrity

²²¹ 06-Essentials 4 Data Support\2. Planning phase\II.3-Reproducible research

²²² 06-Essentials 4 Data Support\4. User phase\IV.5-Data seal of approval

²²³ 09-MANTRA\5 Documentation, metadata, citation\Documentation, metadata, citation23

²²⁴ (GDPR; European Union, 2016a);

²²⁵ 02-CESSDA\5. Protect\5. Protect - CESSDA TRAINING

adequadas, como é feito nos principais repositórios de dados (R02²²⁶). Portanto, ainda que essas precauções tenham sido tomadas no momento do compartilhamento, o pesquisador que reusar dados precisa estar ciente dessas exigências, pois ao reusar dados, este se configura no papel de processador, portanto, está sob os requerimentos legais da proteção de dados e de privacidade.

O pesquisador deve conhecer os repositórios que cumprem esses dispositivos legais. Em alguns casos, o repositório solicita ao controlador de dados (normalmente quando é um órgão público, como depositante dos dados), a indicação do contato do encarregado (Data Protection Officer - DPO), que é cargo responsável pela segurança dos dados. Se necessário, um relatório de avaliação de impacto de privacidade (Data Privacy Impact Assessments - DPIA) pode ser solicitado para esclarecimentos adicionais (R08²²⁷). Uma forma de representação possível é:

CA 12. Reúso (*Reuse*)

UR 12.1 Pesquisa Reprodutível (Reproducible Research)

UC Identificar o contexto em que os dados são produzidos e reutilizados (ciclo de vida dos dados) (F02). Reconhecer que os dados podem ter valor além do objetivo original, para validar pesquisas ou para uso por terceiros (F01). Preparar os dados para reúso de forma ética no futuro. (F06).

UR 12.2 Requisitos Legais (Legal Requirements)

UC Estar ciente das condições para a reutilização de dados (F02). Possui entendimento básico dos requisitos legais em torno da pesquisa (F04).

UR 12.3 Políticas de Reúso (Re-Use Policies)

UC Estar ciente das políticas e finalidades do repositório de dados (F02). Compreender a vida útil do uso potencial de dados de terceiros (F06). Diminuir problemas de versão e duplicação de dados (F06).

- Requisito 13 – **Recompartilhamento** (*Re-sharing*)

O décimo terceiro requisito é o **Recompartilhamento** (*Re-sharing*). Ao considerar o novo compartilhamento dos dados secundários, o pesquisador precisará compreender a política de reutilização desses dados, identificando eventuais restrições (R10²²⁸). Essas possíveis

²²⁶ 02-CESSDA\5. Protect\5.2 Ethical review process - CESSDA TRAINING

²²⁷ 08-Managing and Sharing Research Data handbook\8 RIGHTS RELATING TO RESEARCH DATA\CHAPTER 8 UK Data Service 3

²²⁸ 10-NECDMC\6 Data Sharing E Reuse Policies\necdmc_lessonplan6

restrições legais referem-se à reutilização, se o compartilhamento de dados pode implicar direitos de propriedade de terceiros, incluindo patentes, direitos autorais, direitos de banco de dados, segredos comerciais ou outras informações protegidas por acordos de licença ou não divulgação (*Non Disclosure Agreement* - NDA); direitos do titular dos dados, relacionados ao conteúdo da informação, como privacidade, difamação, etc. (R10²²⁹).

O pesquisador também precisará compreender as políticas de reutilização, de compartilhamento e de acesso aos dados regulados pelas instituições, editores e agências de financiamento às quais a sua própria pesquisa está sujeita (R10²³⁰). Devem ser observadas quais são as permissões da licença atribuída aos dados, se estas permitem o re-compartilhamento. Uma das licenças mais comuns é o conjunto de licenças Creative Commons, que indica quando os dados forem reutilizados, quais são as possíveis opções de compartilhamento de trabalhos derivados (R07²³¹). Na intenção de recompartilhar dados de outro pesquisador, é necessário conferir se há a a permissão para arquivá-los e publicá-los em um repositório de dados (R02²³²). Como representação tem-se:

CA 13. Recompilhamento (*Re-Sharing*)

UR 13.1 Recompilhamento (Re-Sharing)

UC Compreender os benefícios de disponibilizar dados em formatos padrão para facilitar o reuso (F01). Aplicar as melhores práticas para seu futuro compartilhamento de dados (F06).

- Requisito 14 – **Uso dos Dados** (*Using*)

O décimo quarto requisito é o **Uso dos Dados** (*Using*). Essa categoria está dividida em Bases de dados, que são os repositórios de dados, e os conjuntos de dados, que são os dados para reuso, armazenados nos repositórios. Um repositório de dados é uma coleção de arquivo digital que coleta, preserva e exibe conjuntos de dados, documentação e metadados relacionados (OPEN AIRE, 2017²³³) (R02²³⁴). Os repositórios foram projetados para armazenar

²²⁹ 10-NECDMC\6 Data Sharing E Reuse Policies\necdmc_module6

²³⁰ 10-NECDMC\6 Data Sharing E Reuse Policies\necdmc_lessonplan6

²³¹ 07-FOSTER-The Open Data Handbook\07-Full\A FAIRy tale

²³² 02-CESSDA\5. Protect\5.7 Copyright - CESSDA TRAINING

²³³ <https://www.openaire.eu/briefpaper-rdm-infonoads>

²³⁴ 02-CESSDA\6. Archive & Publish\6.3 Data publishing routes - CESSDA TRAINING

a longo prazo (R02²³⁵) e disseminar os dados (R02²³⁶). É permitido embargar um conjunto de dados por um período fixo (até que os autores dos dados publiquem seus resultados, por exemplo) ou impor restrições à reutilização (R01²³⁷).

Uma das maneiras de selecionar dados que não foram inutilizados ao longo do tempo, é escolher um repositório de dados (confiável) que tenha os atributos da curadoria incorporados em sua infraestrutura, que armazene os dados de forma sustentável a manter sua usabilidade (R02²³⁸). Os dados, divulgados a partir dos repositórios, podem ser entregues em diferentes formatos (R02²³⁹). Muitos financiadores e periódicos de pesquisa requerem que os dados sejam disponibilizados em um repositório de dados (R02²⁴⁰). O pesquisador também precisa ser capaz de diferenciar outros serviços de publicação de dados além do repositório, como periódicos de dados ou o auto-arquivamento (R02²⁴¹).

Se não localizar um repositório de dados disciplinar, deve buscar-se um repositório de dados de pesquisa institucional. Alguns editores fornecem indicações de repositórios recomendados (R02²⁴²). Repositórios de dados confiáveis são dedicados a aumentar a capacidade de descoberta de seus conjuntos de dados (R02²⁴³), por isso utilizam um identificador persistente para seus conjuntos de dados, que é automaticamente atribuído no momento do depósito (R02²⁴⁴). A atribuição de identificadores permanentes aos arquivos depositados evita os *Links* mortos e erros "404 não encontrados", o que compromete o avanço da pesquisa (R10²⁴⁵).

Apesar da disponibilidade de diversos repositórios de dados (R02²⁴⁶) e de alguns manterem dados de diversas disciplinas (R07²⁴⁷), o pesquisador deve buscar os dados da pesquisa em um repositório comunitário apropriado ao seu tema (R10²⁴⁸). Algumas universidades têm seus próprios repositórios de dados que oferecem aos pesquisadores a

²³⁵ 02-CESSDA\1. Plan\1.6.1 DMPQuestionsQualitativeData

²³⁶ 02-CESSDA\1. Plan\1.6.2 DMPQuestionsQuantitativeData

²³⁷ 01-Bodleian\Managing research data and data management planning\Research-data-management-planning-introduction-handbook

²³⁸ 02-CESSDA\1. Plan\1.4 FAIR data - CESSDA TRAINING

²³⁹ 02-CESSDA\1. Plan\1.6.2 DMPQuestionsQuantitativeData

²⁴⁰ 02-CESSDA\5. Protect\5.1 Ethics and data protection - CESSDA TRAINING

²⁴¹ 02-CESSDA\6. Archive & Publish\6. Archive & Publish - CESSDA TRAINING

²⁴² 02-CESSDA\6. Archive & Publish\6.3 Data publishing routes - CESSDA TRAINING

²⁴³ 02-CESSDA\6. Archive & Publish\6.7 Access categories - CESSDA TRAINING

²⁴⁴ 02-CESSDA\6. Archive & Publish\6.5 Citing your data - CESSDA TRAINING

²⁴⁵ 10-NECDMC\7 Repositories, Archiving E Preservation\necdmc_module7

²⁴⁶ 02-CESSDA\1. Plan\1.6.1 DMPQuestionsQualitativeData

²⁴⁷ 07-FOSTER-The Open Data Handbook\07-Full\fooster-Data sources _ Find, reuse and cite data _ Library _ University of Leeds

²⁴⁸ 10-NECDMC\6 Data Sharing E Reuse Policies\necdmc_module

facilidade de depositar, compartilhar e licenciar seus recursos de dados para descoberta e uso por outros (R09²⁴⁹). Avaliar criticamente o escopo e as políticas dos repositórios é mandatório (R10²⁵⁰). Ao localizar um repositório confiável, identificar as questões relacionadas à descoberta de conjuntos de dados relevantes para reutilização, compreendendo a necessidade de métodos de busca e recuperação. Há ajuda de várias comunidades de usuários para encontrar os dados que eles desejam que sejam reutilizados (R10²⁵¹).

Os repositórios podem ser certificados por terceiros como repositórios digitais confiáveis (Trusted Digital Repository - TDR), indicando que dados de pesquisa depositados continuarão a ser encontrados e compartilhados no futuro. O Selo de Aprovação de Dados é fundamentado no chamado Modelo de Referência OAIS para um Sistema de Arquivamento Aberto (OAIS Reference Model for an Open Archival System).

Existem, por exemplo, outras formas de certificação, que constituem a estrutura dos repositórios digitais confiáveis europeus (Framework of European Trusted Digital Repositories²⁵²) são: a certificação básica - Selo de aprovação de dados (Data Seal of Approval-DSA²⁵³); a ampla certificação padrão - DIN 31644²⁵⁴ que estabelece critérios para arquivos digitais confiáveis e a certificação formal ISO 16363²⁵⁵, que certifica e audita sistemas de transferência de dados e informações de repositórios digitais confiáveis.

É importante lembrar que as bibliotecas apoiam a criação de coleções digitais, fornecendo repositórios de dados ou assistência na pesquisa em busca de conjuntos de dados (R10²⁵⁶). Muitas instituições têm bibliotecários dedicados a manter seu repositório institucional para coletar os resultados acadêmicos de sua instituição (Teses e Dissertações Eletrônicas, pré-impresões e artigos, além de dados) (R07²⁵⁷) (R10²⁵⁸). Esses bibliotecários também podem ajudar a selecionar um repositório em um campo específico de estudo ou identificar outros recursos disponíveis para dar suporte às suas necessidades de dados de pesquisa (R10²⁵⁹).

Quanto aos conjuntos de dados, os principais são de dados quantitativos ou qualitativos. Dados quantitativos podem se referir apenas às matrizes de números ou palavras que compõem

²⁴⁹ 09-MANTRA\\8 Sharing, preservation & licensing\\Sharing, preservation, and licensing15

²⁵⁰ 10-NECDMC\\7 Repositories, Archiving E Preservation\\necdmc_module7

²⁵¹ 10-NECDMC\\data_management_frameworks\\data_management_frameworks

²⁵² <http://www.trusteddigitalrepository.eu/Welcome.html>

²⁵³ <https://www.datasealofapproval.org/en/>

²⁵⁴ <https://www.din.de/en/getting-involved/standardscommittees/nid/wdc-beuth:din21:147058907>

²⁵⁵ http://www.iso.org/iso/catalogue_detail.htm?csnumber=56510

²⁵⁶ 10-NECDMC\\1 Overview of Research Data Management\\necdmc_module1

²⁵⁷ 07-FOSTER-The Open Data Handbook\\07-Full\\foster_Managing and Sharing Research Data

²⁵⁸ 10-NECDMC\\1 Overview of Research Data Management\\necdmc_module1 (2)

²⁵⁹ 10-NECDMC\\7 Repositories, Archiving E Preservation\\necdmc_module7

um arquivo de dados, mas também podem abranger outras informações (metadados) mantidas em um arquivo de dados estatísticos, como rótulos de variáveis, rótulos de código e definições de valores ausentes ou não, normalmente com uso questionários (R08²⁶⁰). Os dados qualitativos podem incluir transcrições de entrevistas, bem como gravações de áudio e vídeo (analógicas ou digitais) (R08²⁶¹). Os questionários estão marcados com números de identificação (ID) que podem ser vinculados a um número pessoal usando um código de chave, que permite coletar e adicionar dados do registro para cada indivíduo no conjunto de dados. O código da chave permanecerá apenas até que os dados do registro sejam coletados (R02²⁶²). Todas as variáveis de identificação devem ser incluídas no início do seu arquivo de dados geralmente incluem uma identificação exclusiva do seu estudo / arquivo de dados, número único de identificação de casos no seu arquivo de dados (por exemplo, identificação do entrevistado, identificação de sua família etc.), bem como a identificação de outras características essenciais para análise (por exemplo, identificação de diferentes métodos de coleta ou fontes de dados, identificação da amostra excedente, etc.) (R02²⁶³). O relatório de pesquisa apresenta a descrição completa da coleta de dados (R02²⁶⁴). Portanto, o pesquisador que reusa dados precisa tanto dos conjuntos de dados como do relatório de pesquisa, para identificar todas essas relações. Os controles de acesso podem permitir o controle até um nível de arquivo individual, o que significa que níveis mistos de controle de acesso podem ser aplicados a um conjunto de dados (R02²⁶⁵).

O estudo no qual o conjunto de dados está fundamentado deve ter excelência científica e a metodológica deve estar concluída. A relevância para reutilização do conjunto de dados deve ser avaliada por uma ampla variedade de questões práticas e teóricas, como o atendimento aos critérios de qualidade que são determinados com base em extensa documentação acompanhada (R02²⁶⁶). O conjunto de dados deve estar acessível ao público em um repositório de dados científicos, de forma que permita a repetição de descobertas cientificamente publicadas, conduzidas com base na coleta desses dados (R02²⁶⁷).

Portanto, alguns critérios para uso das coleções de dados podem ser identificados: conjuntos de dados acessados e acompanhados de documentação adequada para permitir seu

²⁶⁰ 08-Managing and Sharing Research Data handbook\5 FORMATTING AND ORGANIZING DATA\CHAPTER 5 UK Data Archive - OUR QUALITY CONTROL

²⁶¹ 08-Managing and Sharing Research Data handbook\5 FORMATTING AND ORGANIZING DATA\CHAPTER 5 UK Data Archive - OUR QUALITY CONTROL

²⁶² 02-CESSDA\1. Plan\1.6.2 DMPQuestionsQuantitativeData

²⁶³ 02-CESSDA\3. Process\3.2 Quantitative coding - CESSDA TRAINING

²⁶⁴ 02-CESSDA\1. Plan\1.6.2 DMPQuestionsQuantitativeData

²⁶⁵ 02-CESSDA\5. Protect\5.6 Anonymisation - CESSDA TRAINING

²⁶⁶ 02-CESSDA\6. Archive & Publish\6.4 Publishing with CESSDA archives - CESSDA TRAINING

²⁶⁷ 02-CESSDA\6. Archive & Publish\6.4 Publishing with CESSDA archives - CESSDA TRAINING

uso para fins analíticos e de pesquisa; verificados e validados de acordo com procedimentos rigorosos de coleta de dados; são catalogados de acordo com os padrões de metadados apropriados; os conjuntos de dados, a documentação, os metadados e outras informações de representação são mantidos em condições adequadas para armazenamento de arquivo a longo prazo; são indexados com termos de palavras-chave usando um dicionário de sinônimos apropriado; a autenticidade, a integridade e a confiabilidade das coleções de dados preservados para uso futuro sejam mantidas, quando houver uma nova versão de um conjunto de dados, os metadados descritivos e estruturais relevantes sejam revisados e o arquivo antigo mantido (R08²⁶⁸). A representação proposta:

CA 14. Uso (*Using*)

UR 14.1 Base de Dados (Databases)

UC Compreende o conceito de bancos de dados relacionais, como consultá-los e familiariza-se com os formatos e tipos de dados padrão para sua disciplina (F01).

UR 14.2 Conjunto De Dados (Data Collections)

UC Capacidade de determinar quando os dados são necessários (F02).

- Requisito 15 – **Visualização** (*Visualization*)

O décimo quinto requisito é a **Visualização** (*Visualization*). Não foram localizadas nos recursos de aprendizagem descrições relevantes sobre o processo de visualização dos resultados, sendo mantidas as descrições dos *frameworks* (F01; F02). Recomenda-se a seguinte representação:

CA 15. Visualização (*Visualization*)

UR 15.1 Ferramentas de Visualização (Visualization Tools)

UC Apresentar informações quantitativas (dados específicos, tabelas, gráficos, em relatórios e similares) (F02); Evitar representações enganosas ou ambíguas ao apresentar dados. (F01); Compreender as vantagens de diferentes tipos de visualização, por exemplo, mapas, gráficos, animações ou vídeos, ao exibir

²⁶⁸ 08-Managing and Sharing Research Data handbook\5 FORMATTING AND ORGANIZING DATA\CHAPTER 5 UK Data Archive - OUR QUALITY CONTROL

dados. (F01); Usar proficientemente as ferramentas básicas de visualização da disciplina. (F01).

- Requisito 16 – **Suporte e Treinamento** (*Support and formal training*)

O décimo sexto requisito é o **Suporte e Treinamento** (*Support and formal training*): O letramento em dados pode educar os pesquisadores que não estão familiarizados com o reuso de dados (KIM, 2017). Conhecer e entender a reutilização de dados pode ajudar o pesquisador a (re)modelar todo o seu projeto de pesquisa. Existem as preocupações dos pesquisadores sobre má interpretação e violação na reutilização de dados (tanto pelo uso indevido dos seus dados compartilhados quanto sobre a reutilização de dados de outras pessoas) (KIM, 2016; 2017).

Como essas preocupações podem ser um impedimento importante para reutilizar dados, é importante promover a educação e o apoio institucional para que esses pesquisadores aumentem o nível de confiança dos dados. Fornecer recursos e suportes pode facilitar o reuso de dados pelos pesquisadores, ajudando a minimizar os esforços individuais. Repositórios oferecem suporte para a compreensão do reuso de dados, como guias e ferramentas para ajuda na exploração e manipulação dos dados (R07²⁶⁹), assim como orientação de uso e compreensão de metadados, de documentação e de como os dados podem ser anonimizados (R08²⁷⁰).

As bibliotecas universitárias devem incluir em seus portfólios de serviços o letramento para a aquisição como as fontes de dados existentes, os repositórios, o processamento e os padrões de citação de dados de dados, visto que já é uma prática dessas instituições promover o letramento informacional de outros recursos científicos para a pesquisa (KIM, 2017). Os pesquisadores devem buscar conhecer e programas de capacitação e treinamento das Universidades e instituições sobre Gestão de Dados de Pesquisas (KIM, 2016; KIM, 2017; BERGHMANS *et al.*, 2017). Propõe-se:

CA 16. Treinamento e Suporte (*Support and Formal Training*)

UR 16.1 Treinamento e Suporte (Support and Formal Training)

UC Identificar e localizar os recursos disponíveis para auxílio (*Dataset 1*). Usar as tecnologias de informação, pois são recursos que mais auxiliam os

²⁶⁹ 07-FOSTER-The Open Data Handbook\07-Full\fooster-Reuse data _ Find, reuse and cite data _ Library _ University of Leeds

²⁷⁰ 08-Managing and Sharing Research Data handbook\5 FORMATTING AND ORGANIZING DATA\CHAPTER 5 UK Data Archive - OUR QUALITY CONTROL

pesquisadores. Buscar o apoio de pessoas especializadas nas bibliotecas acadêmicas (*Dataset 2*) Usar as tecnologias de informação que estão disponíveis para suporte. Receber treinamentos com especialistas em gestão de dados (*Dataset 1*) (*Dataset 2*) (*Dataset 3*).

Como forma de validação, esses requisitos e competências foram submetidos, bibliograficamente, às recomendações das melhores práticas da W3C. Todas as 35 melhores práticas da W3C foram identificadas como aplicáveis aos requisitos de reuso de dados (LÓSCIO; BURLE; CALEGARI, 2017). Sendo assim, os requisitos identificados foram comparados com as práticas e apresentados a seguir.

6.5 VALIDAÇÃO DO GRAU DE ADERÊNCIA ÀS RECOMENDAÇÕES W3C

Os requisitos propostos são decorrentes das análises de conteúdo e foram validados a partir do seu grau de aderência às recomendações das melhores práticas em gestão de dados na *internet*, estabelecidas pela W3C (LÓSCIO; BURLE; CALEGARI, 2017). De acordo com Lóscio, Burle e Calegari (2017), recomendar essas práticas ficou a cargo do Web Best Practices Working Group (Grupo de Trabalho das Melhores Práticas), criado para desenvolver o ecossistema de dados abertos, facilitando a comunicação entre desenvolvedores e editores por meio de orientação para melhorar a consistência na maneira como os dados são gerenciados, promovendo a sua reutilização.

Publicado como uma recomendação e sem caráter normativo, as melhores práticas descritas foram desenvolvidas para incentivar e permitir a continuada expansão da Web como meio de troca de dados devido ao crescimento do compartilhamento de dados *online*, como por exemplo: dados abertos de governos em todo o mundo, a crescente publicação *online* de dados de pesquisa incentivados por organizações como a Research Data Alliance (RDA), a coleta, a análise e a publicação *online* de dados de mídia social, o fornecimento de informações em massa e a crescente presença na Web de importantes coleções do patrimônio cultural, como a da Bibliothèque Nationale da França (BNF) (LÓSCIO; BURLE; CALEGARI, 2017).

Portanto, a importância para a padronização na Web se justifica porque esse crescimento não é consistente em estilo e, em muitos casos, não utiliza todo o potencial da capacidade da Web para vincular um fato ao outro, ou para descobrir recursos relacionados e criar visualizações interativas. Ou seja, publicadores dos dados almejam compartilhar os dados, seja

de forma aberta ou com controle de acesso, enquanto os consumidores querem localizar e usar dados, especialmente os precisos, atualizados e disponíveis. Isso fundamenta a necessidade de um entendimento entre publicadores e consumidores de dados, sem o qual os esforços de publicação podem ser incompatíveis com as demandas de uso, que requerem a facilidade de localização e de compreensão dos dados (LÓSCIO; BURLE; CALEGARI, 2017).

O grau de aderência foi determinado pela quantidade de recomendações correspondentes ou relacionadas aos requisitos e competências propostos pela W3C, esta organização de *per se*, registra que existe a possibilidade do estabelecimento de grau de aderência a partir da correspondência com as evidências propostas. Por exemplo, algumas recomendações são similares com mais de uma competência, conforme pode ser visto na relação entre a recomendação “***Provide data license information***” (prover informação sobre licença de dados) com as competências “*Intellectual Property Rights, Licenses types and copyright*”. Todas essas competências referem-se às formas e às regras sobre atribuição de licenças aos conjuntos de dados. Neste caso, tem-se a relação de três competências com uma única recomendação.

Já para outras competências não haviam quaisquer recomendações correspondentes, como por exemplo “*Open Data Journals*”. Não foi localizada nenhuma recomendação correspondente na lista da W3C que se relacione com os periódicos de dados. Portanto, essa seria uma competência não aderente à W3C. Sequencialmente, elencou-se quantos requisitos e competências essenciais não tinham qualquer relação com as recomendações, evidenciando o seu percentual de representatividade dentre os demais - Gráfico 20, retomado mais adiante.

Finalizou-se, assim, o objetivo específico 6 (OE6), com o conjunto de requisitos e competências relacionados com as recomendações da W3C, conforme apresentado no Quadro 28 e no Quadro 29. As recomendações da W3C são apresentadas no Quadro 28. Na análise de aderência, alguns requisitos e competências não encontraram aderência com as recomendações para as melhores práticas da W3C, enquanto outros foram relacionados com mais de uma recomendação, como apresentado no Quadro 29.

Quadro 28 - OE6: Aderência das recomendações das melhores práticas W3C.

Num.	Melhores práticas W3C	Requisitos identificados
01	<i>Provide metadata</i>	<i>Metadata</i>
02	<i>Provide descriptive metadata</i>	<i>Metadata Description</i>
03	<i>Provide structural metadata</i>	<i>Metadata Standards</i>
04	<i>Provide data license information</i>	<i>Intellectual Property Rights, Licenses types and copyright</i>
05	<i>Provide data provenance information</i>	<i>Citation, Quality Assurance</i>
06	<i>Provide data quality information</i>	<i>Data sets description, Quality Assurance</i>
07	<i>Provide a version indicator</i>	<i>Data sets description, Citation</i>
08	<i>Provide version history</i>	<i>Data sets description</i>
09	<i>Use persistent URIs as identifiers of datasets</i>	<i>Citation, Data sets description</i>
10	<i>Use persistent URIs as identifiers within datasets</i>	<i>Citation, Data sets description</i>
11	<i>Assign URIs to dataset versions and series</i>	<i>Citation</i>
12	<i>Use machine-readable standardized data formats</i>	<i>Citation, Metadata Standards</i>
13	<i>Use locale-neutral data representations</i>	
14	<i>Provide data in multiple formats</i>	<i>Formats</i>
15	<i>Reuse vocabularies, preferably standardized ones</i>	<i>Open Data use and reuse</i>
16	<i>Choose the right formalization level</i>	
17	<i>Provide bulk download</i>	<i>Obtaining data, Collecting Data</i>
18	<i>Provide Subsets for Large Datasets</i>	<i>Access data, Naming conventions, Data collections</i>
19	<i>Use content negotiation for serving data available in multiple formats</i>	<i>Processing, Formats</i>
20	<i>Provide real-time access</i>	<i>Access data, Access data source</i>
21	<i>Provide data up to date</i>	<i>Evaluating, Quality Assurance</i>
22	<i>Provide an explanation for data that is not available</i>	<i>Data sets description, Metadata, Description, Quality Assurance</i>
23	<i>Make data available through an API</i>	<i>Access data sources, Databases</i>
24	<i>Use Web Standards as the foundation of APIs</i>	
25	<i>Provide complete documentation for your API</i>	
26	<i>Avoid Breaking Changes to Your API</i>	
27	<i>Preserve identifiers</i>	<i>Citation</i>
28	<i>Assess dataset coverage</i>	<i>Access data</i>
29	<i>Gather feedback from data consumers</i>	<i>Support and formal training</i>
30	<i>Make feedback available</i>	<i>Support and formal training</i>
31	<i>Enrich data by generating new data</i>	<i>Re-sharing, Open Data use and reuse, Reproducible Research</i>
32	<i>Provide Complementary Presentations</i>	<i>Quality Assurance</i>
33	<i>Provide Feedback to the Original Publisher</i>	<i>Publisher rights</i>
34	<i>Follow Licensing Terms</i>	<i>Citation, Intellectual Property Rights, Licenses types and copyright, Publisher rights, Using data ethically</i>
35	<i>Cite the Original Publication</i>	<i>Citation, Using data ethically</i>

Fonte: Adaptado de Lóscio, Burle e Calegari (2017).

Quadro 29 - OE6: Aderência dos requisitos às recomendações das melhores práticas W3C

> 3 recomendações 2 recomendações 1 recomendação

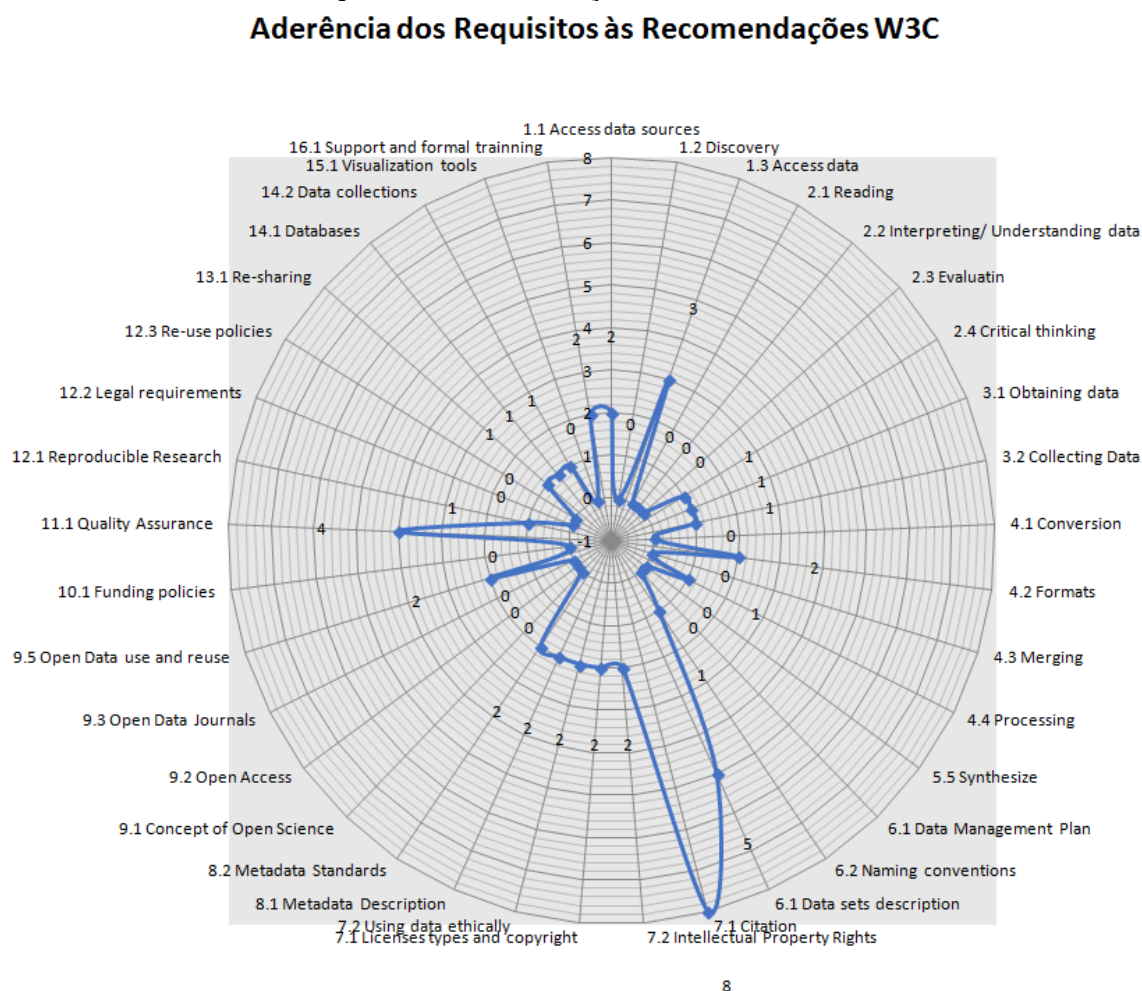
Proposta		Melhores Práticas W3C																																				
Requisitos	Competências	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35		
1. Access	1.1 Access data sources																				X			X														
	1.2 Discovery																																					
	1.3 Access data																			X	X									X								
2. Analysis	2.1 Reading																						X															
	2.2 Interpreting data																						X															
	2.3 Evaluating / 2.4 Critical																						X															
3. Acquisition	3.1 Obtaining data																						X															
	3.2 Collecting Data																						X															
4. Curation	4.1 Conversion																																					
	4.2 Formats															X							X															
	4.3 Merging																																					
	4.4 Processing																						X															
	4.5 Synthesize																																					
5. Data Management	6.1 Data Management Plan																																					
	6.2 Naming conventions																				X																	
6. Documentation	6.1 Data sets description																																					
	7.1 Citation				X	X		X	X			X	X	X											X											X	X	
7. Licences	7.2 Intellectual Property				X																																X	
	7.3 Licenses types and				X																																X	
	7.4 Using data ethically																																		X	X	X	
	8.1 Metadata Description	X	X																																			
8. Metadata	8.2 Metadata Standards	X	X																																			
	9.1 Concept of Open																																					
	9.2 Open Access																																					
	9.3 Open Data Journals																																					
9. Open Science	9.5 Open Data use and																																				X	
	10.1 Funding policies																																					
10. Policy	10.1 Funding policies																																					
	11.1 Quality Assurance						X	X																													X	
11. Quality	12.1 Reproducible Research																																				X	
	12.2 Legal requirements																																					
	12.3 Re-use policies																																					
12. Reuse	13.1 Re-sharing																																				X	
	14.1 Databases																																					
13. Sharing	14.2 Data collections																																					
	15.1 Visualization tools																																					
14. Using	15.1 Visualization tools																																					
	16.1 Support and formal																																				X	X
15. Visualization	16.1 Support and formal																																				X	X
16. Support and formal	16.1 Support and formal																																				X	X

Fonte: Autoria própria (2019).

Dentre os requisitos e as competências recomendados para o reúso, as que não tiveram aderência às recomendações foram: *1.2 Discovery*, *2.1 Reading*, *2.2 Interpreting/ Understanding data*, *5.1 Conversion*, *5.3 Merging*, *5.5 Synthesize*, *6.1 Data Management Plan*, *10.1 Concept of Open Science*, *10.2 Open Access*, *10.3 Open Data Journals*, *11.1 Funding policies*, *13.2 Legal requirements*, *13.3 Re-use policies* e *17.1 Visualization tools*.

Porém, destaca-se que algumas competências tiveram aderência a mais de três recomendações, sendo fortemente ligadas às boas práticas da W3C, são elas: *1.3 Access data*, *7.1 Data sets description*, *8.2 Citation* e *12.1 Quality Assurance*. O nível de aderência dos requisitos às melhores práticas é apresentado no Gráfico 20.

Gráfico 20 - Aderência dos Requisitos às Recomendações W3C



Fonte: Autoria própria (2019).

Nota-se que o requisito 7.1 *Citation* é o que encontra maior aderência junto à W3C, sendo também um dos requisitos mais abordados nos recursos de aprendizado. Quanto aos requisitos sem aderência, como os *1.2 Discovery*, *2.1 Reading*, *2.2 Interpreting/ Understanding*

data, salienta-se que são competências que requerem dos pesquisadores maior capacidade analítica do que técnica, ênfase do conjunto de recomendações da W3C.

Os requisitos *6.1 Data Management Plan*, *10.1 Concept of Open Science*, *10.2 Open Access*, *10.3 Open Data Journals*, *11.1 Funding policies*, *13.2 Legal requirements*, *13.3 Re-use policies*, também não tiveram aderência. Porém, estes são relacionados às competências mais próximas dos conceitos acadêmicos e do reúso de dados de pesquisa, considerando que as recomendações são mais direcionadas para uso de dados de forma mais ampla na *internet*, o que pode justificar a não aderência à W3C.

Já as competências *5.1 Conversion*, *5.3 Merging* e *17.1 Visualization tools* apresentam características de funcionalidades de *softwares* e sob esta ótica apresentam mais prejuízos ao conjunto proposto por não encontrarem aderência às recomendações da W3C. Considera-se, portanto, que apenas essas três competências (*5.1 Conversion*, *5.3 Merging* e *17.1 Visualization tools*), dentre as 37 propostas para reúso, não estejam realmente aderentes às recomendações. Sendo assim, têm-se um percentual de aderência de 92% entre as demais competências e as recomendações da W3C, o que se infere ser um percentual suficiente para validar a proposta. Seguindo esses resultados, são apresentadas as considerações finais e as sugestões para trabalhos futuros.

Esses requisitos e competências propostos constituem-se as diretrizes para a formulação de conteúdos com vistas a compor o programa didático do ambiente em formato de AVA para o reúso de dados. A base para a proposição vem de Pereira, Schmitt e Dias (2007), que registram que as interações dos alunos com o conteúdo no espaço virtual é uma das prerrogativas do processo educativo, o que reforça a importância de conteúdos consistentes. Recorda-se também Mattar (2014), para quem este ambiente deve ter o dinamismo e a personalização da aprendizagem.

Em Pereira, Schmitt e Dias (2007), o conteúdo pode ser escalonado em graus de aprofundamento de acordo com critérios definidos de criação e de seleção e Mattar (2014) salienta que o conteudista e o *designer* instrucional devem considerar as descrições dos requisitos e das competências como base para a formulação dos elementos-chave do AVA, reforçando as proposições ora elencadas.

Esgotadas a apresentação e a análise de dados, e feitas as proposições de requisitos e competências, sequencialmente, apresentam-se as considerações finais.

7 CONSIDERAÇÕES FINAIS

Neste capítulo resgata-se o problema e a pergunta de pesquisa, apresentando a resposta à questão de pesquisa, aos pressupostos iniciais e aos objetivos delimitados. Sequencialmente, são consideradas as limitações da pesquisa e as sugestões para futuras investigações sobre o tema.

7.1 RESPOSTA À QUESTÃO DE PESQUISA E AOS OBJETIVOS

Em vários países do mundo, houve o aumento da cobrança e da exigência pela transparência e prestação de contas como forma de contrapartida ao financiamento público de pesquisas, o que motiva a exigência da disponibilidade de dados de pesquisa em acesso aberto. A problemática desta pesquisa apoiou-se nas potencialidades do reuso desses dados, amplamente disponíveis em um ambiente de pesquisa cada vez mais orientado à Ciência Aberta e pelas exigências de compartilhamento, *versus* a necessidade de competência informacional dos pesquisadores para este reuso, tendo como pano de fundo a Teoria Ator-Rede, que equaliza a agência dos atores envolvidos. A questão de pesquisa delineada diante desse cenário **Quais são os requisitos necessários em ambientes virtuais de aprendizagem para que os pesquisadores desenvolvam competências mínimas de letramento informacional para reuso de dados de pesquisa nas Ciências Sociais, na abordagem da Teoria Ator-Rede?** conduziu as escolhas teóricas e metodológicas do objeto pesquisado.

Verifica-se que o primeiro **Pressuposto (1)**, que considera que a "**sobrecarga de informações do mundo digital se estende à produção científica**", confirmou-se, inclusive sobre próprio tema dos dados de pesquisa. O volume de conteúdo sobre o tema é expressivo, conforme mostra o resultado do levantamento bibliográfico na seção 5.1, sendo, dentre os temas transversais da pesquisa, o conceito-chave mais publicado. Dentre os temas abordados na literatura sobre os dados de pesquisa, há uma abrangência ampla, cobrindo desde a sua geração, o tratamento, a disponibilização e o reuso, sendo esta última a essência deste trabalho.

Esse pressuposto também se confirma, posto que no Conjunto 1 de dados analisados, 86% dos pesquisadores nas Ciências Sociais e 79% nas demais ciências pesquisadas não receberam instrução especializada para os dados de pesquisa. No Conjunto 2 de dados, os números são de 79% dos pesquisadores nas Ciências Sociais e 85% nas demais áreas, respectivamente. E no Conjunto 3 de dados, 53% dos pesquisadores nas Ciências Sociais e 54%

das demais ciências também concordam que não receberam treinamento adequado. Ainda que mais de 50% de pesquisadores de todas as disciplinas reconheçam benefícios e utilidades no reuso de dados, 60% deles acham que é muito esforço reusar dados e não sabem qual licença escolher para isto, além do que 75% dos respondentes não reconhecem metadados, requisito essencial para o reuso de dados. Confirma-se, portanto, o pressuposto de que **"é essencial que os pesquisadores tenham o letramento informacional em dados que lhes permitirá determinar, efetivamente, a credibilidade na localização, no acesso e no reuso de recursos digitais"**, isto a partir dos resultados de opiniões extraídos dos conjuntos de dados secundários.

Recordando, o segundo **Pressuposto (2)**, trata de que a **"biblioteca acadêmica é o locus adequado para promover o letramento da gestão, do compartilhamento e do reuso de dados de pesquisa, considerando seu papel social e de responsabilidade com a geração de conhecimento e não apenas com o depósito do mesmo"**. O pressuposto é confirmado pela revisão de literatura e pelos recursos de aprendizagem. A revisão de literatura demonstrou que diversos serviços de dados têm sido ofertados pelas bibliotecas acadêmicas em vários países. Esses serviços referem-se ao apoio de hospedagem de dados em repositórios, à seleção, à documentação, à citação, ao planejamento de gestão de dados, à propriedade intelectual e direitos autorais, à privacidade, à confidencialidade e preservação e ao letramento em dados, que também compõem os serviços que são de competência das bibliotecas universitárias²⁷¹.

Os Recursos de Aprendizagem também confirmam o pressuposto, indicando que a Biblioteca tem um papel de suporte na gestão dos dados. O recurso 10 indica que "as bibliotecas apoiam a criação de coleções digitais, fornecendo repositórios de dados ou assistência na pesquisa em busca de conjuntos de dados (R10²⁷²). Muitas instituições têm bibliotecários dedicados a manter seu repositório institucional para coletar os resultados acadêmicos de sua instituição"

²⁷¹ Conforme Koltay (2015b, 2017b); Tenopir, Birch e Allard (2012); Tenopir, Hughes e Allard (2015); Mandinach e Gummer (2013); Tenopir *et al.*, (2013, 2014, 2015a, 2016, 2017) e demais autores (ADAMICK; REZNIK-ZELLEN; SHERIDAN, 2012; ASSOCIATION OF COLLEGE AND RESEARCH LIBRARIES, 2014, 2015; BINGERT; BUDDENBOHM, 2016; BROWN, 2013; CARLSON *et al.*, 2011; CARLSON; JOHNSTON, 2015; CREASER, 2011; CHRISTENSEN-DALSGAARD *et al.*, 2012; CULLEN; CHAWNER, 2011; DECHMAN; SYMS, 2014; DELANEY; BATES, 2015; ELLWEIN *et al.*, 2014; FEDERER *et al.*, 2016; FONG; WANG, 2015; FRANK; PHARO, 2016; HELBIG, 2016; JAHNKE; ASHER; KERALIS, 2012; JEFFRYES; JOHNSTON, 2013; JOHNSON; BRESNAHAN, 2015; KARDELIENE; KARDELIS; BAKUTYTE, 2014; KOLTAY; HORSTMANN; WITT, 2016; MACY; COATES, 2016; MANDINACH; GUMMER, 2013; MAYERNIK *et al.*, 2014; MASSACHUSETTS INSTITUTE OF TECHNOLOGY, 2014; MOONEY *et al.*, 2014; QIN; D'IGNAZIO, 2010; SURKIS *et al.*, 2017; UNIVERSITY OF MINNESOTA, 2014; VERBAAN; COX, 2014; WANNER, 2015; WITT, 2012; WHITMIRE, 2015).

²⁷² 10-NECDMC\1 Overview of Research Data Management\necdmc_module1

Os Recursos de Aprendizagem *Online* também são desenvolvidos, mantidos e até financiados por bibliotecas de universidades e de pesquisa, como:

- R01 – Bodleian Data Library – University of Oxford; disponibilizado pela Biblioteca Bodleian da Universidade de Oxford; o R03 - Data Management Course (resultado do DIL Project), um trabalho conjunto das bibliotecas da Universidade Purdue, da Universidade Cornell, da Universidade de Minnesota e da Universidade de Oregon, este financiado por Institute of Museum and Library Science;
- R09–MANTRA Research Data and Management Training, mantido pelos profissionais de apoio e suporte aos serviços de dados, da biblioteca da Universidade de Edimburgo;
- R10 – New England Collaborative Data Management Curriculum (NECDMC) que é liderado por uma coalizão de bibliotecários da Biblioteca Lamar Soutter da Escola de Medicina da Universidade de Massachusetts em parceria com várias bibliotecas da região da Nova Inglaterra;
- R11 – Research Data Management and Sharing - Coursera, desenvolvido pelo Curating Research Assets and Data Using Lifecycle Education (CRADLE), em colaboração com a Biblioteca da Universidade da Carolina do Norte em Chapel Hill e com EDINA em Universidade de Edimburgo. O terceiro pressuposto (3), registra que:

[...] os dados de pesquisa que eram produzidos e esquecidos, desorganizados e não acessíveis nos arquivos pessoais dos pesquisadores ou das instituições, na medida em que são tratados, padronizados, preservados e disponíveis de forma gratuita, constituem um novo actante para se fazer Ciência, a partir do seu reuso. (ESTEVÃO, 2019, p. 34).

Trata-se, portanto, da disponibilidade dos dados, plenamente confirmada. Não apenas o referencial teórico sobre os dados de pesquisa é expressivo, como também os conjuntos de dados disponíveis por meio de repositórios o são. O indexador re3data²⁷³, lista, em setembro de 2019, 1.796 repositórios de dados de pesquisa, em todas as áreas do conhecimento. Cada repositório hospeda milhares de conjuntos de dados, e o pesquisador está consciente dessa disponibilidade, como mostra os resultados das análises dos dados secundários. Há um consenso positivo de 60% em torno do acesso aos repositórios e que dados estão disponíveis, além de que entre 60% e 90% consideram a disponibilidade como algo valioso, desejável e útil às pesquisas. Portanto, confirma-se o pressuposto de que os dados estão sendo padronizados e disponibilizados, configurando um novo actante para se fazer Ciência.

²⁷³ <http://re3data.org/>

Isto posto, resgatam-se os objetivos da pesquisa para demonstrar se essas assertivas foram atendidas. O objetivo geral de "Propor, a partir de elementos da Teoria Ator-Rede, requisitos para um ambiente virtual em letramento informacional de dados, com as competências mínimas para o reúso de dados em Ciências Sociais" foi cumprido e está detalhado no Capítulo 6 ao longo da Seção 6.4. Esse resultado é, em especial, uma descrição detalhada dos requisitos e das competências identificadas, nos moldes das recomendações de Latour (2000), de forma a instruir o pesquisador não habituado ao reúso de dados, sobre as implicações de que este não saberia antecipadamente como reusar os dados até se deparar com a necessidade ou escolha do reúso de dados. Para o cumprimento deste objetivo, seis objetivos específicos foram delimitados, tendo sido todos atingidos e discutidos em sequência.

O Objetivo Específico 1 – OE1 previu a caracterização dos ambientes virtuais de pesquisa formados pela infraestrutura de coleta, de armazenamento, de serviços, de disponibilidade e de reúso de dados de pesquisa com base na literatura. O resultado alcançado é a contextualização descrita nos Capítulos 2 e 3, evidenciando as práticas relacionadas ao gerenciamento de dados de pesquisa nos ambientes, por meio de serviços de curadoria e de infraestrutura institucional destinada aos pesquisadores, impulsionados pela ciência orientada à dados. Inevitavelmente, o processo da comunicação científica também foi modificado, na medida em que os dados passaram a intervir na dinâmica da rede de disseminação do conhecimento científico. Sob as lentes da Teoria Ator-Rede, a agência dos atores na rede de comunicação científica pode ser observado de maneira simétrica entre elementos humanos e não humanos, e como esses interferem nos arranjos sócio técnicos sob os quais a Ciência se desenvolve. A ciência aberta, o acesso aberto e os dados abertos promovem uma abordagem mais participativa e cidadã da Ciência, evidenciando princípios democráticos e de interesse público da produção do conhecimento.

O Objetivo Específico 2 – OE2, previu o levantamento dos quadros conceituais (*frameworks*) de letramento informacional em dados, que visavam fundamentar os requisitos propostos para o reúso de dados. Os principais *frameworks* utilizados (CARLSON *et al.*, 2011; CALZADA PRADO; MARZAL, 2013; CARLSON; JOHNSTON, 2015; JEFFRYES; JOHNSTON, 2013; FOSTER, 2015; VITAE, 2011) foram analisados na perspectiva da Análise de Conteúdo. Os resultados desse objetivo estão no Capítulo 5, na Seção 5.9 e especificamente no Quadro 19, após o resultados das análises. Também há conteúdos complementares que detalham os *frameworks* nos Apêndices G, H, I, J e K.

Este levantamento demonstrou algumas homogeneidades entre os requisitos propostos, que mesmo usando diferentes terminologias, abordam conceitos como localizar,

descobrir, obter, avaliar, gerenciar, organizar, processar, usar, converter, atribuir metadados, documentar, preservar, reusar, respeitar questões legais e éticas, garantir a qualidade e visualizar dados. Todos esses resultados de estudos são respaldados com dados empíricos de pesquisadores que estavam sujeitos aos dispositivos regulatórios de compartilhamento de dados em outros países. Os *frameworks* evidenciaram a necessidade de equalizar os requisitos para a gestão de dados, uma nova instância com a qual os pesquisadores precisam lidar.

O Objetivo Específico 3 – OE3 identificou as perspectivas e as dificuldades dos pesquisadores no reuso dados de pesquisa, por meio da utilização de dados secundários (*datasets*). Foram utilizados os conjuntos de dados de Kim (2016[*datasets*];), os de Kim, (2017 [*datasets*]) e os de Berghmans *et al.* (2017, [*datasets*]), ordenados no Capítulo 6, Seção 6.1, especialmente de forma comparada no Quadro 21. Nesta análise, algumas competências requeridas diante das dificuldades apresentadas pelos pesquisadores já foram identificadas. A análise demonstrou que os pesquisadores tinham receios com o uso inadequado ou equivocado dos conjuntos de dados; consideram um grande esforço localizar os dados e melhorar a qualidade da sua documentação; precisam de compreensão mais aprofundada das regras de propriedade intelectual, das regras de direitos autorais e de licenças dos dados, das condições para reconhecer a qualidade dos dados para o reuso e, principalmente, esperam ações de capacitação e treinamento das instituições às quais estão vinculados, o que demonstra a aderência deste trabalho aos seus propósitos.

O Objetivo Específico 4 – OE4, identificou as competências essenciais para o reuso de dados no letramento proposto pelos recursos de aprendizado *online*, disponíveis nos ambientes virtuais de aprendizagem. Esse resultado pode ser encontrado no Capítulo 6, Seção 6.2 no Quadro 24, em que as competências foram classificadas por recorrência e relevância a partir das estratégias de pesquisa. As competências mais recorrentes indicam relações com **regras de acesso, tipos de licença e como usar os conjuntos de dados** disponíveis. Ainda que o tema dados abertos de pesquisa seja intrinsecamente ligado à Ciência Aberta, que pressupõe o mínimo de barreiras, o fato de um dos resultados mais recorrentes estar ligado às questões de licenças de direitos autorais demonstra como a forma de recompensa da Academia se impõe como regra de compartilhamento.

O Objetivo Específico 5 – OE5, comparou as características dos quadros conceituais com as perspectivas dos pesquisadores e com as competências nos recursos de aprendizagem *online*, propondo um conjunto de requisitos com vistas ao reuso de dados nas Ciências Sociais. Esse conjunto de requisitos está disponível no Capítulo 6, Seção 6.3 e no Quadro 25. Essa

comparação foi importante para evidenciar intersecções de competências e de requisitos disponíveis nas três instâncias de análise.

O Objetivo Específico 6 – OE6, validou os requisitos propostos por meio da verificação de aderência às melhores práticas da área recomendadas pelo World Wide Web Consortium (W3C) e consta no Capítulo 6, Seção 6.5, no conjunto de requisitos, conforme apresentado no Quadro 28. Essa validação, no entanto, apesar de ter usado um modelo de referência, não foi suficiente para validar todos os requisitos, o que pode ser explicado da seguinte forma: das 4 competências mais aderentes à W3C, pelo critério de atendimento à mais de três recomendações (*1.3 Access data*, *7.1 Data sets description*, *8.2 Citation* e *12.1 Quality Assurance*) duas delas estão relacionadas com os requisitos mais recorrentes nos recursos de aprendizagem, ou seja, o Acesso (*1.3 Access data*) e a Licença (*8.2 Citation*). Isso demonstra o forte alinhamento entre os recursos de aprendizagem e as melhores práticas, evidenciando que os temas mais recorrentes abordados nos AVAs estão presentes em mais de três recomendações da W3C. Ou seja, são requisitos imprescindíveis quando se trata de reuso de dados disponíveis *online*.

Ademais, as outras duas competências mais aderentes (*7.1 Data sets description* e *12.1 Quality Assurance*) constituem-se essenciais para o cumprimento do terceiro requisito mais recorrente nos recursos de aprendizagem, que é o uso efetivo dos conjuntos de dados. Sem uma boa documentação (que se refere as descrições dos conjuntos de dados), que permite o entendimento do contexto dos dados, e a garantia da qualidade, atestada pela proveniência e relevância dos mesmos, o reuso dos conjuntos de dados se torna inviável. Desta forma, ainda que as recomendações da W3C não tenham sido suficientes para validar todas as competências, já que as competências *5.1 Conversion*, *5.3 Merging* e *17.1 Visualization tools* não foram possíveis de serem relacionadas com as recomendações, os requisitos mais relevantes, no que tange ao reuso, foram todos fortemente aderentes ao modelo de validação, com mais de três correlações: Acesso, Licenças e Uso.

Neste trabalho, defendeu-se a tese de que **ambientes virtuais de aprendizagem podem ser recursos de suporte à Pesquisa e ampliar a possibilidade de reuso de dados se ofertarem o letramento informacional aos pesquisadores a partir de um conjunto de requisitos aplicáveis ao reuso, principalmente os critérios que atestem a proveniência, a qualidade e a confiabilidade dos dados secundários.**

Neste contexto, com base na literatura e nas instâncias de análises, como os *frameworks*, os dados secundários e os recursos de aprendizagem, infere-se que os ambientes virtuais de pesquisa tornam-se mais integrados, viabilizados pela evolução das tecnologias de comunicação e da informação, ao mesmo tempo em que a própria Ciência utiliza tais inovações,

modificando a forma como os cientistas trabalham e compartilham os seus resultados. Sob a égide da Ciência Aberta, outras formas de colaboração foram sendo ampliadas, buscando não apenas uma racionalização dos investimentos, mas principalmente a abertura de diversas práticas que disponibilizam recursos a fim de promover o desenvolvimento da Ciência, como os dados de pesquisa, aproximando-se de um uso sustentável, inclusive dos dados de pesquisa.

Em contrapartida, habilidades técnicas e conhecimento jurídico podem constituir barreiras ao pesquisador na apropriação dos benefícios potenciais deste novo paradigma, impulsionados pela abordagem de abertura e colaboração do ambiente científico e se beneficiar da disponibilidade de dados para reuso. Ou seja, a capacidade de aproveitamento dos recursos gerados a partir da prática da Ciência Aberta, como os dados de pesquisa disponíveis em acesso aberto, pode estar limitada pela falta destes conhecimentos. A incerteza em torno do conteúdo disponível também limita o pesquisador, que receia ter seus resultados comprometidos ou cometer alguma infração.

Neste sentido, a proposta de requisitos para o letramento no reuso de dados abertos pode contribuir assertiva e qualitativamente para a atividade fim do pesquisador, que é produzir Ciência, e comprova a tese, pois apresenta as competências e as habilidades para direcioná-lo no acesso, na apropriação, no uso efetivo desses recursos, reconhecendo os critérios de qualidade e de confiabilidade, de acordo com as permissões legais, dentre outras habilidades e competências propostas no conjuntos de requisitos, disponível no Quadro 26.

Isso contribui para um ciclo virtuoso da comunicação dos resultados das pesquisas, em que mais ativos digitais são utilizados com menos barreiras possíveis, tendo os seus resultados também disponibilizados com o mínimo de restrições. Estabelece-se, portanto, um processo de retro alimentação, colaboração e integração, suportado por recursos tecnológicos, direcionados por uma abordagem mais democrática dos recursos científicos, promovendo o compartilhamento do conhecimento, ao mesmo tempo em que este processo está adequado à regulação aplicável. A identificação de 92% de aderência desses requisitos e competências às recomendações de melhores práticas propostas pela W3C reforça a comprovação da tese, principalmente porque todas as 35 práticas recomendadas viabilizam o reuso de dados, objeto desta pesquisa e principal elemento da tese.

7.2 LIMITAÇÕES DA PESQUISA

Para a elaboração deste trabalho foram utilizados recursos da literatura, *locus* da coleta dos *frameworks*, além dos dados secundários e dos recursos de aprendizagem *online*,

recuperados por meio de indexadores de repositórios de dados. As estratégias de busca precisaram ser testadas e refinadas nas bases de dados na busca pela literatura, o que exigiu conhecimento especializado, sendo esta uma barreira a ser vencida, conforme atestam os dados analisados e a sua prática de uso.

A metodologia de análise de conteúdo foi utilizada nas três instâncias de análise, e, por ser exaustiva, gera muitos resultados, o que demanda o manejo de vários arquivos separados. O volume de análise dos recursos de aprendizagem, somando os 11 cursos, totalizou 695 arquivos entre documentos *pdf*, *sites*, vídeos e apresentações. Foi necessário um *software* de apoio para a análise qualitativa, neste caso selecionado o NVivo, tanto para a organização dos arquivos, como para o processamento das pesquisas, a extração e organização dos resultados. Muitas horas foram necessárias para apropriação de saberes sobre a ferramenta utilizada – o letramento da própria pesquisadora, entre treinamentos *online*, tutoriais e vídeos. O processo de aplicação dos conhecimentos adquiridos, testes, resultados negativos, erros e correções levou aproximadamente 6 meses.

O reuso de dados secundários configurou-se como um desafio. Essa é uma prática não usual e não há protocolos ou recomendações prontas. Algumas competências, como a busca, a localização, a obtenção, a descoberta, a curadoria (limpeza de dados, filtros, anonimização dos dados, transformação de arquivos), a avaliação de parâmetros de qualidade, a citação e a seleção de conteúdo, tiveram de ser exercidas antes das análises dos recursos de aprendizagem, o que exigiu mais uma etapa de capacitação para a pesquisa. O Plano de Gestão de dados também não é um requisito exigido nos programas de pós-graduação, portanto, não há referências nacionais sobre o tema. A construção do plano levou em conta quatro modelos mais recomendados por agências de financiamentos internacionais. Grande parte dos recursos sobre o tema de dados de pesquisa estão em inglês, o que exige a proficiência no idioma.

Apesar de serem estas limitações comuns às pesquisas de modo geral, neste estudo, o desafio foi também o de registrar adequadamente todos os dados sob o enfoque da Ciência Aberta e, infere-se, sustentável, além de adaptada aos vieses filosóficos do Programa de Pós-Graduação em Tecnologia e Sociedade, multidisciplinar, da Linha de Pesquisa Tecnologia e Desenvolvimento, do Grupo de Pesquisa e da Orientação.

7.3 SUGESTÕES PARA TRABALHOS FUTUROS E ÚLTIMAS CONSIDERAÇÕES

Os resultados da pesquisa em tela podem ser considerados um ponto de partida para a habilitação de outros pesquisadores no reuso de dados, ainda sendo necessárias futuras investigações. A continuação da presente pesquisa pode expandir os conjuntos de requisitos, contextualizados com outras práticas de gestão de dados, como o compartilhamento, a reprodutibilidade e o escrutínio de resultados publicados, as aplicações tecnológicas e os *softwares* que automatizem os processos de análises de dados, e ainda, os que melhorem a visualização dos resultados obtidos por meio dessas análises.

O aprofundamento dos requisitos e das competências em letramento para o reuso também deve ser explorado, incluindo formas dinâmicas de avaliação de viabilidade da adesão ao reuso de dados de acordo com critérios da investigação pretendida, como em áreas e disciplinas específicas, objetivos, metodologia, cobertura, finalidade ou atualidade diversas.

Pesquisas futuras também deverão considerar fatores de níveis políticos e organizacionais que influenciam o reuso dos dados, como políticas e exigências de órgãos de fomento para uma determinada área do conhecimento ou tipo de pesquisa, ou ainda pesquisas financiadas com recursos públicos e a relação proporcional de disponibilidade de dados nesta área.

Outro aspecto de investigação futura necessária é a mensuração econômica e financeira do reuso de dados em projetos de ampla coleta e que utilizem dados públicos, visto que muitos destes dados já estão disponíveis por meio da legislação de dados abertos dos governos ou disponibilizados por força da lei de acesso a informação (LAI), no contexto brasileiro.

Percebe-se, no âmbito da legislação, que este tema deve ser aprofundado sob a ótica da privacidade de dados pessoais. Ainda é imprecisa a forma como a gestão dos dados custodiados por pesquisadores e universidades pode ser impactada por legislações de privacidade como a Regulamento Geral sobre a Proteção de Dados (GDPR) no âmbito Europeu, as regulamentações norte americanas sobre proteção e privacidade de dados (CCPA, HIPAA, FERPA, PII²⁷⁴) e a Lei Geral de Proteção de Dados (LGPD), no Brasil, principalmente pelos riscos de falhas na anonimização dos dados de pesquisa.

Esses estudos podem investigar questões de equilíbrio entre proteção da privacidade e do acesso aberto, segurança da informação, governança de dados e riscos cibernéticos para a

²⁷⁴ California Consumer Protection Act (CCPA); Health Insurance Portability and Accountability Act of 1996 (HIPAA); Family Educational Rights and Privacy Act of 1974 (FERPA); Personally Identifiable Information (PII).

Instituição e para os repositórios de dados, com foco na avaliação de riscos e responsabilidades associados a quaisquer violações dos, e incidentes com, dados.

Constatou-se que, efetivamente, os pesquisadores não estão habituados a reutilizar dados de pesquisa no contexto da ciência aberta, como mostra a análise dos dados secundários e demonstram estar pouco preparados para esta tarefa, evidenciando a relevância dos achados desta pesquisa, em que os principais requisitos como o Acesso, as Licenças e o Uso, abrangem desde etapas iniciais até as finais do reúso de dados. Esses requisitos cobrem desde as competências preliminares, como o acesso – ponto de partida para viabilidade do reúso, passando pelas permissões a serem observadas, até o efetivo uso das coleções de dados, que envolvem, dentre outras ações, a manipulação destes. Dessa forma, o letramento para o reúso de dados passa pelo letramento informacional, como o acesso, a descoberta, a avaliação crítica e os direitos autorais de conteúdo dos *datasets*, como o letramento em dados em si, as atividades de curadoria, de limpeza de dados e da manipulação de arquivos eletrônicos e sistemas relacionados.

Em última análise, a emergência dos AVAs pode ser compreendida com um processo sociotécnico, pois valeu-se dos avanços científicos, das tecnologias de informação e de comunicação (TICs) viabilizadas pela internet e por demandas sociais que, assim como em outros ambientes, transcendem barreiras geográficas e temporais para a prática da educação à distância, corroborando-se com a literatura pesquisada e utilizada²⁷⁵. Como características dos processos sociotécnicos, os ambientes virtuais educativos para o reúso de dados também estarão permeados de interações e inter-relação complexas entre atores²⁷⁶, como os pesquisadores que irão interagir com ferramentas concebidas para o ensino a distância, tendo esta tecnologia não como um fim em si mesma, antes como uma plataforma que permite a ação. Esta ação visa o aprendizado dessas competências para o reúso dos conjuntos de dados, que por sua vez estão carregados de regras institucionais e legislativas sobre direitos autorais, normas acadêmicas e modelos metodológicos sob os quais foram construídos, arquiteturas computacionais, lógica e linguagens de programação dos sistemas que ditam como os arquivos eletrônicos deverão ser manipulados, salvos, armazenados e posteriormente recuperados.

Um desses requisitos é a elaboração do Plano de Gestão de Dados (*Data Management Plan* – DMP), recurso previsto também para esta pesquisa e que foi elaborado de forma voluntária e está disponível no Apêndice D. Este é um registro importante, pois considera a tratativa a ser aplicada a todos os conteúdos e dados cujas análises fundamentam os resultados

²⁷⁵ Neste caso específico com Pereira, Schmitt e Dias (2007).

²⁷⁶ (CALLON; LAW, 1989; LINSINGEN; BAZZO; PEREIRA, 2003; MATTAR, 2014)

desta pesquisa. Também se ressalta a sua relevância, pois compõe a documentação desta pesquisa, auxiliando em futuras análises, escrutínio ou reuso dos dados ora apresentados.

As interações no ensino à distância ocorrem com os atores humanos entre si (alunos com instrutores, alunos com alunos) e entre actantes humanos e não humanos (alunos com o conteúdo, alunos com a ferramenta, alunos com a interface)²⁷⁷, e a inter-relação de atores heterogêneos nos AVAs, que são constituídos de suas mídias e conteúdo (elaborados por diversos profissionais) e que intermediam o processo junto aos alunos, promovem os movimentos de integração e de colaboração²⁷⁸, aproximando-se do que Latour (2001) denomina de mediação técnica.

Neste caso, da mediação técnica do ensino-aprendizagem à distância para o reuso de dados, em que os actantes não humanos também possuem agência, destaca-se que os requisitos e as competências, achados teóricos desta pesquisa, possuem agência em si mesmos, pois ao serem incorporados ou apropriados por meio da interação no ambiente virtual, consolidam a aprendizagem²⁷⁹; os usuários terão vantagens no desenvolvimento de suas pesquisas, podendo fazer uso de maior variedade de fontes e de recursos de dados, distintos do que tinham antes, ainda que estes dados já estivessem disponíveis. Contudo, salienta-se que apenas a disponibilidade não gera possibilidade de reutilização, sendo essencial o letramento para reuso, com o desenvolvimento de competências como: acessar, obter e reusar os dados, respeitando os direitos autorais e de propriedade intelectual. Dessa forma, os pesquisadores serão impulsionados a usufruir dos benefícios proporcionados pelos recursos da Ciência Aberta e especialmente pelas vantagens que os dados secundários representam.

²⁷⁷ (MATTAR, 2014)

²⁷⁸ (PEREIRA; SCHMITT; DIAS, 2007)

²⁷⁹ (MATTAR, 2014)

REFERÊNCIAS

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **NBR 10520**: Citações em documentos. Rio de Janeiro: ABNT, 2002. 7 p.

ADAMICK, Jessica; REZNIK-ZELLEN, Rebecca; SHERIDAN, Matt. Data Management Training for Graduate Students at a Large Research University. **Journal of eScience Librarianship**, Boston, v. 1, n. 3, p. 180–188, 2013. University of Massachusetts Medical School. Disponível em: <http://dx.doi.org/10.7191/jeslib.2012.1022>. Acesso em: 21 jul. 2018.

ALBAGLI, Sarita; Ciência aberta em questão. *In*: ALBAGLI, Sarita; MACIEL, Maria Lucia; ABDO, Alexandre Hannud. **Ciência aberta, questões abertas**. Brasília: IBICT; Rio de Janeiro: UNIRIO, 2015. p. 9-26.

ALBAGLI, Sarita. Território e Territorialidade. *In*: LAGES, Vinícius; BRAGA, Christiano; MORELLI, Gustavo (Org.). **Territórios em movimento**: cultura e identidade como estratégia de inserção competitiva. Rio de Janeiro: Relume Dumará / Brasília: Sebrae, 2004. p. 23-70.

AMERICAN LIBRARY ASSOCIATION (ALA). **Presidential Commission on Information Literacy Final Report**. American Library Association: Chicago, IL, 1989. 6 p.

AMORIM, Ricardo Carvalho; CASTRO, João Aguiar; SILVA, João Rocha da; *et al.* A Comparative Study of Platforms for Research Data Management: Interoperability, Metadata Capabilities and Integration Potential. *In*: ROCHA, Álvaro; CORREIA, Ana Maria; COSTANZO, Sandra; *et al* (Org.). **New Contributions in Information Systems and Technologies**. Switzerland: Springer International Publishing, 2015, v. 353, p. 101–111. Disponível em: http://link.springer.com/10.1007/978-3-319-16486-1_10. Acesso em: 21 mar. 2018.

ANDRADE, Rodrigo de Oliveira. Disseminação desigual: Reúso de informações científicas ainda é baixo e varia de acordo com a área do conhecimento. **Pesquisa Fapesp**, São Paulo, n. 265, mar. 2018, p.44-47.

ASSOCIATION OF COLLEGE AND RESEARCH LIBRARIES (ACRL). ACRL Information Literacy Best Practices Committee. Characteristics of programs of information literacy that illustrate best practices: A guideline Approved by the ACRL Board of Directors, June 2003, **College & Research Libraries News**, v. 73 n. 6, 2012a, p. 355-359. Disponível em: <https://crln.acrl.org/index.php/crlnews/article/view/8781/9349>. Acesso em: 21 jan. 2018.

ASSOCIATION OF COLLEGE AND RESEARCH LIBRARIES (ACRL). ACRL Research Planning and Review Committee. Top ten trends in academic libraries. A review of the trends and issues affecting academic libraries in higher education, **College & Research Libraries News**, v. 73 n. 6, 2012b, p. 311-320. Disponível em: <https://goo.gl/x81bQY>. Acesso em: 21 jan. 2018.

ASSOCIATION OF COLLEGE AND RESEARCH LIBRARIES (ACRL). ACRL Research Planning and Review Committee. Top ten trends in academic libraries. A review of the trends and issues affecting academic libraries in higher education. **College & Research Libraries**

News, v.6, n. 75, 2014, p. 294–302. Disponível em: <https://goo.gl/L2DJWy>. Acesso em: 21 jan. 2018.

ASSOCIATION OF COLLEGE AND RESEARCH LIBRARIES (ACRL). **Framework for Information Literacy for Higher Education**. Association of College and Research Libraries: Chicago, IL: 2015. Disponível em <http://www.ala.org/acrl/standards/ilframework>. Acesso em: 21 jan. 2018.

ASSOCIATION OF COLLEGE AND RESEARCH LIBRARIES (ACRL). **Information Literacy Competency Standards for Higher Education**. Association of College and Research Libraries, Chicago, IL. 2000. Disponível em: <http://www.ala.org/acrl/standards/informationliteracycompetences>. Acesso em: 20 jul. 2017.

ASSOCIATION OF COLLEGE AND RESEARCH LIBRARIES (ACRL). **Intersections of scholarly communication and information literacy: Creating strategic collaborations for a changing academic environment**. Chicago, IL: Association of College and Research Libraries. 2013. Disponível em: <https://goo.gl/cYQfoL>. Acesso em: 21 jan. 2018.

ASSOCIATION OF RESEARCH LIBRARIES (ARL). **About**. [s. d.]. *Online*. Disponível em: <https://goo.gl/Y6ZcTw>. Acesso em: 28 fev. 2017.

ASSOCIATION OF RESEARCH LIBRARIES (ARL). **Monograph & Serial Costs in ARL Libraries, 1986-2011***. 2012. Disponível em: <http://goo.gl/WMQwvC>. Acesso em: 20 jul. 2017.

ASSOCIATION OF RESEARCH LIBRARIES (ARL). **E-Science and data support services: A study of ARL member institutions**. Washington, DC: Association of Research Libraries. 2010. Disponível em: <https://goo.gl/QvYzlu>. Acesso em: 28 fev. 2017.

ASSOCIATION OF RESEARCH LIBRARIES (ARL). **To Stand the Test of Time: Long-Term Stewardship of Digital Data Sets in Science and Engineering. A Report to the National Science Foundation from the ARL Workshop on New Collaborative Relationships—The Role of Academic Libraries in the Digital Data Universe**, Arlington, Virginia, 26-27 set. 2006. **Association of Research Libraries**, 2006. Disponível em: <https://goo.gl/fg9ER8>. Acesso em: 28 fev. 2017.

AVENTURIER, Pascal; ALENCAR, Maria Cleofas de. Os desafios de dados de pesquisa abertos. **Revista Eletrônica de Comunicação, Informação e Inovação em Saúde**, [s.l.], v. 10, n. 3, set. 2016. Disponível em: <https://www.reciis.icict.fiocruz.br/index.php/reciis/article/view/1069>. Acesso em: 08 ago. 2017. doi:<http://dx.doi.org/10.29397/reciis.v10i3.1069>.

AWRE, Chris; BAXTER, Jim; CLIFFORD, Brian; COLCLOUGH, Janette; COX, Andrew; DODS, Nick; DRUMMOND, Paul *et al.* Research Data Management as a “wicked problem”. **Library Review**, [s.l.], v. 64, n. 4/5, p. 356–371, 6 jul. 2015. Disponível em: <https://doi.org/10.1108/LR-04-2015-0043>. Acesso em: 12 mar. 2018.

BALL, Alex. **A review of data management lifecycle models (version 1.0)**. REDm-MED Project Document redm1rep120110ab10. 2012. Bath, United Kingdom: University of Bath. 15p. Disponível em: <https://goo.gl/DoUFRM>. Acesso em: 12 mar. 2018.

BARDIN, Laurence. **Análise de conteúdo**. Lisboa: ed.70, 1995.

BAWDEN, David; ROBINSON, Lyn. The dark side of information: overload, anxiety and other paradoxes and pathologies. **Journal of Information Science**, [s.l.], v. 35, n. 2, p. 180-191. 2009. Disponível em: <https://goo.gl/fgy4hx>. Acesso em: 12 mar. 2018. <https://doi.org/10.1177/0165551508095781>

BERGHMANS, Stephane; COUSIJN, Helena; DEAKIN, Gemma, DEAKIN, Gemma; MEIJER, Ingeborg; MULLIGAN, Adrian; PLUME, Andrew; RIJCKE, Sarah de; RUSHFORTH, Alex; TATUM, Clifford; LEEUWEN, Thed van; WALTMAN, Ludo. **Open Data: the researcher perspective - survey and case studies**. [Dataset]. 2017. Mendeley Data, versão 1, 4 abr. 2017. Disponível em: <https://data.mendeley.com/datasets/bwrnfb4bvh/1>. Acesso em: 25 abr. 2017. DOI: 10.17632/bwrnfb4bvh.1

BERLIN Declaration on open access to knowledge in the sciences and humanities. Berlin, 2003. Disponível em: https://openaccess.mpg.de/67605/berlin_declaration_engl.pdf. Acesso em: 08 set. 2017.

BERMAN, E. Transforming Information Literacy in The Sciences Through the lens of E-Science. **Communications in Information Literacy**, Tulsa, v. 7, n. 2, p. 161-170, 2013. Disponível em: <http://www.comminfolit.org/index.php?journal=cil&page=article&op=viewArticle&path%5B%5D=v7i2p161>. Acesso em: 12 ago. 2017.

BILL & MELINDA GATES FOUNDATION. **Bill & Melinda Gates Foundation open access policy**. 2015. Disponível em: <https://goo.gl/H6CUnH>. Acesso em: 11 ago. 2017.

BINGERT, S., BUDDENBOHM, S. Research data center services for complex software environments in the humanities. **Information Services and Use**, Clifton, v. 36, n. 3-4, 2016, p. 189-202. Disponível em: doi: <https://doi.org/10.3233/ISU-160817>. Acesso em: 21 jan. 2018.

BORGMAN, Christine. L. Data Scholarship in the Social Sciences. *In*: BORGMAN, Christine. L. **Big Data, Little Data, No Data: Scholarship in the Networked World**. Christine L. Borgman. Cambridge, MA: MIT Press, 2015. p. 107-133.

BORGMAN, Christine. L. Digital libraries and the continuum of scholarly Communication. **Journal of Documentation**, London, v. 56, n. 4, jul. 2000, p. 412-430.

BORGMAN, Christine L. Scholarship at a Crossroads. *In*: **Scholarship in the Digital Age: Information, Infrastructure, and the Internet**, 2007a, p. 1-12. Cambridge: MIT Press, 2007a. Disponível em: <http://www.jstor.org/stable/j.ctt5hhbk7.6>.

BORGMAN, Christine L. The Continuity of Scholarly Communication. *In*: **Scholarship in the Digital Age: Information, Infrastructure, and the Internet**, 2007b, p. 47-74. Cambridge: MIT Press, 2007b. Disponível em: <http://www.jstor.org/stable/j.ctt5hhbk7.9>.

BORGMAN, Christine. L. The conundrum of sharing research data. **Journal of the American Society for Information Science and Technology**, North Carolina, 2012, v. 63, n. 6, p.

1059–1078. <https://doi.org/10.1002/asi.22634>.

BORGMAN, Christine, DARCH, Peter, SANDS, Ashley; PASQUETTO, Irene; GOLSHAN, Milena; WALLIS, Jillian; TRAWEEK, Sharon. Knowledge infrastructures in science: data, diversity, and digital libraries. **International Journal on Digital Libraries**, Berlin, v. 16, n. 3/4, p. 207-227, 2015. DOI 10.1007/s00799-015-0157-z.

BORGMAN, Christine L; FURNER, Jonathan. Scholarly communication and bibliometrics. **Annual Review of Information Science and Technology**, v. 36, n. 1, fev. 2005, p. 2–72. Wiley. <https://doi.org/10.1002/aris.1440360102>.

BOSCH, Stephen; HENDERSON, Katie. Coping with the Terrible Twins: Periodicals Price Survey 2012. **Library Journal**. [s.l.], 30 abr. 2012. Disponível em: <http://goo.gl/Yv8Jb>. Acesso em: 20 jul. 2016.

BRASIL. CÂMARA DOS DEPUTADOS. **Projeto de lei N.º 1.120-B, de 2007**. 2007. Disponível em:

https://www.camara.leg.br/proposicoesWeb/prop_mostrarintegra;jsessionid=F86D3B71F1E4FCA4F9135827EBF4063D.node2?codteor=572586&filename=Avulso+-PL+1120/2007. Acesso em: 30 jun. 2017.

BRASIL. CÂMARA DOS DEPUTADOS. Comissão de Constituição e Justiça e de Cidadania (CCJC). 2009. **Apresentação do Parecer do Relator, PRL 1 CCJC, pelo Dep. Valtenir Pereira, em 29 de outubro de 2009**. Disponível em:

https://www.camara.leg.br/proposicoesWeb/prop_mostrarintegra;jsessionid=04C28E520F7EC6AD656A7B261A48CDEA.proposicoesWebExterno2?codteor=707945&filename=Tramitacao-PL+1120/2007. Acesso em: 30 jun. 2017.

BRASIL. CÂMARA DOS DEPUTADOS. **Projetos de Leis e Outras Proposições: PL 1120/2007**. 2012a. Disponível em:

https://www.camara.leg.br/proposicoesWeb/prop_mostrarintegra?codteor=1015640&filename=Tramitacao-PL+1120/2007. Acesso em: 30 jun. 2017.

BRASIL. CÂMARA DOS DEPUTADOS. **Resolução nº 17, de 1989**: regimento interno da Câmara dos Deputados. 2012b. 9. ed. Brasília, 177. Disponível em:

<https://www2.camara.leg.br/atividade-legislativa/legislacao/regimento-interno-da-camara-dos-deputados/arquivos-1/RICD%20atualizado%20ate%20RCD%206-2019.pdf>. Acesso em: 30 jun. 2017.

BRASIL. CÂMARA DOS DEPUTADOS. **Projetos de Leis e Outras Proposições: PL 6702/2013**. 2013a. Disponível em: <http://goo.gl/CWYo33>. Acesso em: 30 jun. 2017.

BRASIL. CÂMARA DOS DEPUTADOS. **Projeto de Lei n. 6.702, de 05 de novembro de 2013**. 2013b. Disponível em: <http://goo.gl/fh4HPR>. Acesso em: 30 jun. 2017.

BRASIL. CÂMARA DOS DEPUTADOS. Comissão de Ciência e Tecnologia, Comunicação e Informática (CCTCI). **Parecer ao Projeto de Lei n. 6.702, de 06 de maio de 2014**. 2014a. Disponível em: <http://goo.gl/7LEFxR>. Acesso em: 30 jun. 2017.

BRASIL. MINISTÉRIO DA EDUCAÇÃO (MEC). COORDENAÇÃO DE APERFEIÇOAMENTO DE PESSOAL DE NÍVEL SUPERIOR (Capes). 2017. **Legislação**. Disponível em: <https://goo.gl/wwwvAFP>. Acesso em: 11 ago. 2017.

BRASIL. MINISTÉRIO DA EDUCAÇÃO (MEC). COORDENAÇÃO DE APERFEIÇOAMENTO DE PESSOAL DE NÍVEL SUPERIOR (Capes). **Orçamentos: Dotação e Execução - PPA 2012-2013.2013d**. Disponível em: <http://goo.gl/MdK11Q>. Acesso em: 27 jun. 2017.

BRASIL. MINISTÉRIO DA EDUCAÇÃO (MEC). COORDENAÇÃO DE APERFEIÇOAMENTO DE PESSOAL DE NÍVEL SUPERIOR (Capes). 2015. **Portal de Periódicos Capes/MEC: Missão e Objetivos**. Acesso em 15 set. 2017. Disponível em: https://www-periodicos-capes-gov-br.ez22.periodicos.capes.gov.br/index.php?option=com_pcontent&view=pcontent&alias=mis-sao-objetivos&Itemid=109&. Acesso em: 21 jan. 2017.

BRASIL. SENADO FEDERAL. **Texto inicial - Projeto de Lei do Senado nº 387, de 2011**.2011. Disponível em: <http://goo.gl/dZTP4E>. Acesso em: 25 jul. 2017.

BRASIL. SENADO FEDERAL. Comissão de Constituição, Justiça e Cidadania (CCJC). **Parecer ao Projeto de Lei do Senado nº 387, de 2011**.2013c. Disponível em: <https://legis.senado.leg.br/sdleg-getter/documento?dm=3777715&ts=1553267702884&disposition=inline>. Acesso em: 25 jul. 2017.

BRASIL. Lei nº 12.527, de 18 de novembro de 2011. Regula o acesso a informações previsto no inciso XXXIII do art. 5º, no inciso II do § 3º do art. 37 e no § 2º do art. 216 da Constituição Federal; altera a Lei no 8.112, de 11 de dezembro de 1990; revoga a Lei no 11.111, de 5 de maio de 2005, e dispositivos da Lei no 8.159, de 8 de janeiro de 1991; e dá outras providências. **Diário Oficial [da] República Federativa do Brasil**, Brasília, DF, 18 dez. 2011. Disponível em: <https://goo.gl/33YHS>. Acesso em: 11 ago. 2017.

BRAZILIAN POLITICAL SCIENCE REVIEW (BPSR). **Author Guidelines**, São Paulo. Disponível em: <https://brazilianpoliticalsciencereview.org/author-guidelines/>. Acesso em: 11 jul. 2017.

BRINEY, Kristin, GOBEN, Abigail, ZILINSKI, Lisa. Do You Have an Institutional Data Policy? A Review of the Current Landscape of Library Data Services and Institutional Data Policies. **Journal of Librarianship & Scholarly Communication**, Forest Grove, v.3, n.2, p. 1–25, 2015. <https://doi.org/10.7710/2162-3309.1232>

BROWN, Mark L. The role of the research library. *In*: JUBB, Michael; SHORLEY, Debbie. **The Future of Scholarly Communication**. London: Facet, 2013. p. 157-168.

BRUCE, Christine S.; HUGHES, Hilary. Informed learning: A pedagogical construct for information literacy. **Library and Information Science Research**, [s.l.], v.32, n.4, A2–A8, 2010. doi:10.1016/j.lisr.2010.07.013.

BRUCE, Christine S.; HUGHES, Hilary; SOMERVILLE, Mary. Supporting informed learners in the 21st century. **Library Trends**, [s.l.], v. 61, n. 3, p. 522–545, 2012.

BRUCE, Christine. S., SOMERVILLE, Mary; STOODLEY, Ian.; PARTRIDGE, Helen. Diversifying information literacy research: An informed learning perspective. *In*: HEPWORTH, M.; WALTON, G. (Eds.). **Developing people's information capabilities: Fostering information literacy in educational, workplace and community contexts**. Bingley, England: Emerald Group Publishing Limited. 2013. p. 225–242.
<https://scholarlycommons.pacific.edu/libraries-books/9>.

BRYANT, Rebecca; LAVOIE, Brian; MALPAS, Constance. **A Tour of the Research Data Management (RDM) Service Space**. The Realities of Research Data Management, Part 1. Dublin, Ohio: OCLC Research. 2017. Disponível em:
<https://www.oclc.org/content/dam/research/publications/2017/oclcresearch-research-data-management-service-space-tour-2017-a4.pdf>. Acesso em: 11 ago. 2017.

BUENO DE LA FUENTE, Gema. **Challenges and strategies for the success of Open Science**. *In*: FACILITATE OPEN SCIENCE TRAINING FOR EUROPEAN RESEARCH (FOSTER). Open Science at the Core of Libraries. 2016a. Disponível em:
<https://www.fosteropenscience.eu/node/1424>. Acesso em 30 set. 2016.

CALLON, Michel. Society in the Making: The Study of Technology as a Tool for Sociological Analysis. *In*: PINCH, Trevor J.; BIJKER, Wiebe E.; HUGHES, Thomas P. **The social Construction of Technological Systems: new directions in the sociology and History of Tecnology**. Cambridge (Mass.), MIT, 1993, p. 77-97. Disponível em:
<https://bibliodiarq.files.wordpress.com/2015/09/bijker-w-the-social-construction-of-technological-systems.pdf>. Acesso em: 15 mar. 2019.

CALLON, Michel. Some elements of a sociology of translation: domestication of the scallops and the fishermen of St Brieuc Bay. *In*: LAW, John. **Power, action and belief: a new sociology of knowledge?** London, Routledge, 1986, p. 196-223.

CALLON, Michel; LAW, John; RIP, A. How to study the force of science. *In*: CALLON, M.; LAW, J.; RIP, A. (Eds.) **Mapping the dynamics of science and technology**. Houndmills: Macmillan Press, 1986, p. 3-15.

CALLON, Michel; LAW, John. On the Construction of Sociotechnical Networks: Content and Context Revisited In JONES, R. A; HARGENS, L.; PICKERING, A. **Knowledge and Society**, v. 8: Studies in the Sociology of Science Past and Present. Greenwich, CT: JAI Press, 1989, p. 57–83.

CALZADA PRADO, Javier; MARZAL, Miguel Ángel. Incorporating Data Literacy into Information Literacy Programs: Core Competencies and Contents. **Libri: International Journal of Libraries & Information Services**, Berlin, v. 63, n.2, p. 123–34. 2013. Disponível em: <https://doi.org/10.1515/libri-2013-0010>. Acesso em: 02 jul. 2017.

CAMPELLO, Bernadete. O movimento da competência informacional: uma perspectiva para o letramento informacional. **Ciência da Informação**, Brasília, v. 32, n. 3, p. 28-37, dez. 2003. Disponível em: <https://goo.gl/8W55Nm>. Acesso em: 14 abr. 2018.
<http://dx.doi.org/10.1590/S0100-19652003000300004>. Acesso em: 02 jan. 2018.

CANADA. GOUVERNEMENT DU QUÉBEC. MINISTÈRE DE L'ÉDUCATION NATIONALE DE LA RECHERCHE ET DE LA TECHNOLOGIE. **Les compétences transversales**. 2004. Programme de formation de l'école québécoise. 2004. 24 p. Disponível em:
http://www.education.gouv.qc.ca/fileadmin/site_web/documents/education/jeunes/pfeq/PFEQ_competences-transversales-deuxieme-cycle-secondaire.pdf. Acesso em: 02 jan. 2018.
 Copyright <http://www.droitauteur.gouv.qc.ca/copyright.php>.

CAREGNATO, Sonia Elisa. O desenvolvimento de habilidades informacionais: o papel das bibliotecas universitárias no contexto da informação digital em rede. **Revista de Biblioteconomia e Comunicação**, Porto Alegre, v. 8, p. 47-55, 2000. Disponível em:
<http://eprints.rclis.org/11663/1/artigoRBC.pdf> /. Acesso em: 02 jul. 2017.

CARLSON, Jake R.; FOSMIRE, Michael; MILLER, Chris; NELSON, Megan R. Sapp. Determining Data Information Literacy Needs: A Study of Students and Research Faculty. **Libraries Faculty and Staff Scholarship and Research**, West Lafayette, Paper 23, p. 1-30, 2011. Disponível em: http://docs.lib.purdue.edu/lib_fsdocs/23. Acesso em: 12 dez. 2017.

CARLSON, Jake; JOHNSTON, Lisa R. **Data Information Literacy: Librarians, Data, and the Education of a New Generation of Researchers**. Purdue Information Literacy Handbooks. West Lafayette, Indiana: Purdue University Press, 2015. 271 p. Disponível em:
<https://goo.gl/riC4Es>. Acesso em: 12 dez. 2017.

CARTER, Ian M. Changing institutional research strategies. *In*: JUBB, Michael; SHORLEY, Debbie. **The Future of Scholarly Communication**. London: Facet, 2013. p. 145-156.

CARVALHO, José; RODRIGUES, Eloy; PRÍNCIPE, Pedro. Projeto FOSTER: facilitar a formação e a disseminação das práticas de Ciência Aberta na Europa, 2014. *In* **V CONFERÊNCIA LUSO-BRASILEIRA SOBRE ACESSO ABERTO**, Coimbra, Portugal, 06-08 de outubro de 2014. Disponível em:
http://repositorium.sdum.uminho.pt/bitstream/1822/30495/2/ConfOA2014_Poster_FOSTER_FINAL.pdf. Acesso em: 12 dez. 2017.

CARVALHO, Teila Oliveira. A Influência das Revistas Científicas de Acesso Aberto para o Depósito e Publicação dos Dados de Pesquisa. *In*: **7. CONFERÊNCIA LUSO-BRASILEIRA SOBRE ACESSO ABERTO. Anais...** 2016. Disponível em: <https://goo.gl/kjZhCM>. Acesso em: 11 ago. 2017.

CENDI. **Implementation of Public Access Programs in Federal Agencies**. 2016. Disponível em: <https://goo.gl/WhJ8Az>. Acesso em: 10 fev. 2018.

CESAR JUNIOR, Roberto Marcondes. Do mundo aos dados e dos dados ao conhecimento. *In*: HEY, Tony; TANSLEY, Stewart; TOLLE, Kristin (org.). **O quarto paradigma: descobertas científicas na era da eScience**. São Paulo: Oficina do Texto, 2011. P. 7-10.
https://web.icmc.usp.br/SCATUSU/Boletim_aquisicao/Boletim_Agosto_2017/Capas_Agosto_2017/Hey_O_quarto0001.pdf

CHRISTENSEN-DALSGAARD, Birte; BERG, Marc van den; GRIM, Rob; HORSTMANN, Wolfram; JANSEN, Dafne; POLLARD, Tom; ROOS, Annikki. **Ten recommendations for libraries to get started with research data management**: Final report of the LIBER working group on E-Science / Research Data Management. Ligue des Bibliothèques Européennes de Recherche, 2012. <https://libereurope.eu/wp-content/uploads/The%20research%20data%20group%202012%20v7%20final.pdf>

CONSORTIUM OF EUROPEAN SOCIAL SCIENCE DATA ARCHIVES (CESSDA). CESSDA Training Working Group. 2018. Data in the social sciences. *In*: **CESSDA Data Management Expert Guide**. Bergen, Norway: CESSDA ERIC. Disponível em: <https://www.cessda.eu/DMGuide>. Acesso em: 23 mar. 2019.

COPE, Bill; KALANTZIZ, Mary. Changing knowledge ecologies and the transformation of the scholarly journal. *In*: COPE, Bill; PHILIPS, Angus. (Eds) **The Future of the Academic Journal**. 2 ed. Oxford: Chandos, 2014. p. 09-74.

CORTI, Louise; EYNDEN, Veerle Van den; BISHOP, Libby; WOOLLARD, Matthew. **Managing and Sharing Research Data**: a Guide to Good Practice, Essex: UK Data Service, 2014. Disponível em: <https://www.ukdataservice.ac.uk/manage-data/handbook>. Acesso em: 20 fev. 2019.

COUNCIL ON LIBRARY AND INFORMATION RESOURCES. **No brief candle**: Reconceiving research libraries for the 21st century. Washington DC: Council on Library and Information Resources, 2008. 81 p. Disponível em: <https://goo.gl/tRLXCR>. Acesso em: 28 fev. 2017.

COUNCIL ON LIBRARY AND INFORMATION RESOURCES. **Research data management**: Principles, practices, and prospects. Council on Library and Information Resources, 2013. 108 p. Disponível em: <https://goo.gl/tZ1N9F>. Acesso em: 12 mar. 2017

CREASER, Claire. Scholarly communication and access to research output. *In*: EVANS, WENDY, BAKER, DAVIS. **Libraries and Society**: role, responsibility and future in an age of change. Boston: EBSCO Publishing, 2011. p. 53-66.

CULLEN, Rowena; CHAWNER, Brenda. Institutional repositories, open access, and scholarly communication: a study of conflicting paradigms. **The Journal of Academic Librarianship**, [s.l.], v. 37, n., 6, 2011, p. 460-470.

CURATING RESEARCH ASSETS AND DATA USING LIFECYCLE EDUCATION (CRADLE); UNIVERSITY OF NORTH CAROLINA; UNIVERSITY OF EDINBURG. **Research Data Management and Sharing**. 2016. Disponível em: <https://www.coursera.org/learn/data-management>. Acesso em: 20 fev. 2019.

CURTY, Renata Gonçalves. As diferentes dimensões do reuso de dados científicos. *In*: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 17., 2016, Bahia. **Anais....** Bahia: Universidade Federal da Bahia, 2016a. p. 1 - 23. Disponível em: <http://www.ufpb.br/evento/index.php/enancib2016/enancib2016/paper/viewFile/4056/2495>. Acesso em: 15 mar. 2019.

CURTY, Renata Gonçalves. **Beyond Data Thrifting**: An Investigation of Factors Influencing

Research Data Reuse in the Social Sciences. 2015. 308 fl. Thesis (Doctor of Philosophy) School of Information Studies, Syracuse University, Syracuse, New York. Disponível em: <https://surface.syr.edu/etd/266>. Acesso em: 20 fev. 2019.

CURTY, Renata Gonçalves. Factors Influencing Research Data Reuse in the Social Sciences: An Exploratory Study. **International Journal of Digital Curation**, Edinburgh, v. 11, n. 1, p. 96-117, 2016b. DOI: <https://doi.org/10.2218/ijdc.v11i1.401>.

CURTY, Renata Gonçalves; CROWSTON, Kevin; SPECHT, Alison; W. GRANT, Bruce; DALTON, Elizabeth D. Attitudes and norms affecting scientists' data reuse. **PloS one**, v.12, n. 12, 2017, e0189288. Disponível em: <https://doi.org/10.1371/journal.pone.0189288>. Acesso em: 02 jan. 2018.

CURTY, Renata Gonçalves; KIM, Youngseek; QIN, Jian. **What have Scientists Planned for Data Sharing and Reuse?** A Content Analysis of NSF Awardees' Data Management Plans. Baltimore: Research Data Access & Preservation Summit (RDAP13), 2013. 27 p.

CUTCLIFFE, Stephen H. La Emergência histórica de CTS como campo acadêmico. *In*: CUTCLIFFE, Stephen H. **Ideas, máquinas y valores: los estudios de Ciência, Tecnología y Sociedad**. Barcelona: Antropos, 2003. p. 7-23.

DATAONE education modules. DataONE. 2012. Disponível em: <https://www.dataone.org/education-modules>. Acesso em: 02 jul. 2018.

DAVIS, P. M., LEWENSTEIN, B.V.; SIMON, D.H.; BOOTH, J.G.; CONNOLLY, M.J.L. Open Access Publishing, Article Downloads, and Citations: randomised controlled trial, **The BMJ**, [s.l.], n. 337, 2008,a568. doi: <https://doi.org/10.1136/bmj.a568>

DECHMAN, Margaret K.; SYMS, Laura R. 2014. Working Together to Maximize the Utilization of Open Data Across Social Science and Professional Disciplines. **Behavioral & Social Sciences Librarian**, [s.l.], v. 33, n. 4, p. 188–207. Disponível em: <https://doi.org/10.1080/01639269.2014.964617>. Acesso em: 02 jan. 2018.

DELANEY, Geraldine; BATES, Jessica. Envisioning the Academic Library: A Reflection on Roles, Relevancy and Relationships. **New Review of Academic Librarianship**, , [s.l.], v. 21, n. 1, p. 30–51, 2015. DOI: 10.1080/13614533.2014.911194.

DIGITAL CURATION CENTER. **DC 101 materials: Access and Reuse**. 2004-2019. Disponível em: <http://www.dcc.ac.uk/training/train-the-trainer/dc-101-training-materials>. Acesso em: 31 jan. 2019.

DIGITAL SCIENCE; FIGSHARE. **Digital Science Report - The State of Open Data: A selection of analyses and articles about open data**. Out. 2016.53 p. Disponível em: <https://dx.doi.org/10.6084/m9.figshare.4036398>. Acesso em: 02 jan. 2018.

DOCUMENTATION AND LIBRARY DIVISION OF THE FACULTY OF SCIENCES AND TECHNOLOGY (Lisboa). New UniversityofLisbon. **Blimunda Project**. Disponível em: <https://sites.google.com/site/blimundaproject/sobre>. Acesso em: 26 fev. 2018.

DULCINEA. *Online*. Disponível em: <https://goo.gl/ueMzLJ>. Acesso em: 26 fev. 2018.

DURAND, Thomas. Forms of incompetence. *In: INTERNATIONAL CONFERENCE ON COMPETENCE-BASED MANAGEMENT*, 4., 1998, Oslo. **Proceedings...** Oslo : Norwegian School of Management, 1998.

DURAND, Thomas. Forms of incompetence. *In: SANCHEZ, Ron; HENEE, Aimée (Ed.). Theory Development for competence-based management*, Greenwich, v. 6, Advanced Business Strategy, Lawrence Foster, series editor, Greenwich, CT: JAI Press, 2000.

ELLWEIN, Amy L.; HARTLEY, Laurel M.; DONOVAN, Sam; BILLICK, Ian. Using Rich Context and Data Exploration to Improve Engagement with Climate Data and Data Literacy: Bringing a Field Station into the College Classroom. **Journal of Geoscience Education**, London, v. 62, n.4, p. 578–586, 2014. <https://doi.org/10.5408/13-034>.

ESTEVIÃO, Janete Saldanha Bach. **Letramento informacional para reúso de dados nas Ciências Sociais em Ambientes virtuais de pesquisa**: proposta de requisitos e competências. 2019. 430 f. Tese (Doutorado em Tecnologia e Sociedade) – Programa de Pós-Graduação em Tecnologia e Sociedade, Universidade Tecnológica Federal do Paraná, Curitiba, 2019.

ESTEVIÃO, Janete Saldanha Bach; ARNS, Elaine Mandelli; STRAUHS, Faimara do Rocio. Gestão de dados de Pesquisa: uma prática para abrir a caixa preta da pesquisa científica. **Rev. Dig. Bibliotecon. Cienc. Inf.**, Campinas, v.17, p. 1-26, 2019.

EUROPEAN COMMISSION. DIRECTORATE-GENERAL FOR RESEARCH AND INNOVATION. **Open innovation, Open Science, open to the world. A vision for Europe**. Brussels: European Commission, 2006c. 108 p. Disponível em: <http://bookshop.europa.eu/en/open-innovation-open-science-open-to-the-world-pbKI0416263>. Acesso em: 20 nov. 2016. DOI: 10.2777/061652.

EUROPEAN COMMISSION. **Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020**. 15 nov. 2016b. 11 p. Disponível em: http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf. Acesso em: 20 jul. 2016.

EUROPEAN COMMISSION. **H2020 Programme. H2020 AGA – Annotated Model Grant Agreement**. Version 2.1.1. 1 jul. 2016a. p. 216-219. Disponível em: <http://goo.gl/SXbKxE>. Acesso em: 20 nov. 2016.

EUROPEAN COMMISSION. **HORIZON 2020 em breves palavras**: O programa-quadro de investigação e inovação da UE. 2014. 40 p. Disponível em: <https://goo.gl/M72BgB>. Acesso em: 20 nov. 2016.

EUROPEAN COMMISSION. Open Science Skills Working Group Report. **Providing researchers with the skills and competencies they need to practice Open Science**. Jul. 2017. Disponível em: https://ec.europa.eu/research/openscience/pdf/os_skills_wgreport_final.pdf#view=fit&pagemode=none. Acesso em: 02 jul. 2019.

EUROPEAN COMMISSION. **Study on the economic and technical evolution of the scientific publication markets in Europe**: final report. Bruxelles: European Commission,

jan. 2006. 112 p. Disponível em:

https://ec.europa.eu/research/openscience/pdf/openaccess/librarians_2006_scientific_publication.pdf. Acesso em: 20 nov. 2016.

EYNDEN, Veerle Van den; SCHRECKENBERG, Kate; CORTI, Louise. Sharing social data in multidisciplinary, multi-stakeholder research. Best practice guide for researchers. **ESPA Working Paper Series**, Edinburg, n. 1, Dez. 2015, Ecosystems Services for Poverty Alleviation programme1 (ESPA); K Data Service (UKDS). 23 p. Disponível em <https://goo.gl/Sxuc1g>. Acesso em 19 jul. 2018.

FACILITATE OPEN SCIENCE TRAINING FOR EUROPEAN RESEARCH. Open Science Learning Objectives, 2015. FOSTER Consortium Members. Disponível em: <https://goo.gl/Lkm1XT>. Acesso em: 20 jul.2016. p. 536-552. Doi: 10.5281/zenodo.15603.

FANIEL, Ixchel. M.; BARRERA-GOMEZ, Julianna; KRIESBERG, Adam; YAKEL, Elizabeth. A Comparative Study of Data Reuse Among Quantitative Social Scientists and Archaeologists. *In*: Conference Fort Worth, Fort Worth, TX, USA, 12-15 fev. 2013, p. 797-800, **Anais....** 2013. <https://core.ac.uk/download/pdf/10201965.pdf>

FANIEL, Ixchel M; ZIMMERMAN, Ann. Beyond the data deluge: A research agenda for largescale data sharing and reuse. **International Journal of Digital Curation**, Edimburgo, v.6, n.1, p. 58-69, 2011. Disponível em: <http://dx.doi.org/10.1787/5jrs2f963zs1-en>. Acesso em: 10 ago. 2017.

FECHER, Benedikt. FRIESIKE, Sascha Open Science: One Term Five Schools of Thoughts. *In* BARTLING, Sönke; FRIESIKE, Sascha (Org). **Opening Science**, p. 17–47. Cham: Springer International Publishing, 2014. https://doi.org/10.1007/978-3-319-00026-8_2. Disponível em: <https://goo.gl/3qrg2o>. Acesso em 28 Jan. 2017.

FEDERER, Lisa M., LU, Ya-Ling; JOUBERT, Douglas J. Data literacy training needs of biomedical researchers. **Journal of the Medical Library Association**, Bethesda, v. 104, n 1, p. 52–57. 2016. <https://doi.org/10.3163/1536-5050.104.1.008>.

FEENBERG, Andrew. Racionalização Subversiva: Tecnologia, Poder e Democracia. *In*: NEDER, Ricardo T. **A teoria crítica de Andrew Feenberg**: racionalização democrática, poder e Tecnologia. Brasília: Observatório do Movimento pela Tecnologia Social na América Latina / CDS / UnB / Capes, 2010, p. 67-95. Disponível em: <https://www.sfu.ca/~andrewf/coletanea.pdf>. Acesso em: 20 nov.2017.

FEENBERG, Andrew. Technologyan freedom: Democratizing Technical Change. *In*: **Alternative Modernity**. The Technical Turn in Philosophy and social Theory. Berkeley/Los Angeles: University of California Press, 1995. P. 7-17. Disponível em: https://www.sfu.ca/~andrewf/books/Alternative_Modernity.pdf. Acesso em: 20 nov.2017. DOI: 10.2307/1400309.

FLORES, Murilo. **A identidade cultural do território como base de estratégias de desenvolvimento**: uma visão do estado da arte. 2006. 47 p. Disponível em: <https://goo.gl/UrPUZC>. Acesso em: 23 fev. 2018.

- FONG, Bonnie, L.; WANG, Minglu. Required Data Management Training for Graduate Students in an Earth and Environmental Sciences Department. **Journal of eScience Librarianship**, Boston, v. 4, n.1, 2015. 11 p. Disponível em: <https://doi.org/10.7191/jeslib.2015.1067>. Acesso em: 02 jan. 2018.
- FRANK, Emily P.; PHARO, Nils. Academic Librarians in Data Information Literacy Instruction: A Case Study in Meteorology. **College & Research Libraries**, Chicago, v. 77, n. 4, Jul. 2016, p. 536-552. Doi:10.5860/crl.77.4.536.
- FRANKEL, Mark S., ELLIOTT, Roger, BLUME, Martin, BOURGOIS, Jean-Manuel, HUGENHOLTZ, Bernt, LINDQUIST, Mats G., MORRIS, Sally, e SANDEWALL, Erik. Defining and certifying electronic publication in science. **Learned Publishing**, [s.l.], v.13, n. 4, 2000, p. 251–258. Disponível em: <https://goo.gl/5y2Mq7>. Acesso em 28 jan. 2017.
- FUNDAÇÃO DE AMPARO À PESQUISA DO ESTADO DE SÃO PAULO. **Plano de Gestão de Dados - FAPESP**. 2018. Disponível em: <http://www.fapesp.br/gestaodedados/>. Acesso em: 04 abr. 2018.
- GIARLO, Michael J. Academic Libraries as Data Quality Hubs. **Journal of Librarianship and Scholarly Communication**, [s.l.], v. 1, n. 3, p.1-10, 2013. Pacific University Library. <http://dx.doi.org/10.7710/2162-3309.1059>.
- GIL, Antônio Carlos. **Como elaborar projetos de pesquisa**. 5. ed. São Paulo: Atlas, 2010.
- GITBOOK BOT, HELLER, Lambert, DATAWOMANHUB, MCANCELLIERI, Bianca Kramer, ROSS-HELLAUER, Tony, TENNANT, Jon. **Open Science Training Handbook** (Version 1.1). Zenodo. <http://doi.org/10.5281/zenodo.1212538>. 2018. Disponível em: <https://book.fosteropenscience.eu/en/>. Acesso em: 02 jul. 2019.
- GODOY, Arlida Schmidt. Introdução à Pesquisa Qualitativa e suas Possibilidades. **Revista de Administração de Empresas**, São Paulo, v. 35, n. 2, p. 57-63, mar./abr. 1995. <http://dx.doi.org/10.1590/S0034-75901995000200008>.
- GRAY, Jim; LIU, David T.; NIETO-SANTISTEBAN, Maria; SZALAY, Alex; DEWITT, David J.; HEBER, Gerd. Scientific data management in the coming decade. **ACM SIGMOD Record**, [s.l.], v. 34, n. 4, p. 34-41, jan. 2005. Disponível em: <https://arxiv.org/ftp/cs/papers/0502/0502008.pdf>. Acesso em: 23 jan. 2018.
- GRIGG, Karen Stanley. Data in the Sciences. In: KELLAM, Lynda; THOMPSON, Kristi, **Datalibrarianship: The Academic Data Librarian in Theory and Practice**. Association of College and Research Libraries. Chicago: American Library Association, 2016, p. 179-192.
- GUEDON, Jean-Claude. **Why are electronic publications difficult to classify:**The orthogonality of print & digital media.1994. 7 p. Disponível em: <http://www.people.virginia.edu/~pm9k/libsci/guedon.html>. Acesso em: 27 jan. 2018.
- GUERTIN, Hélène. BERNHARD, Paulette. **Les 6 étapes d'un projet de recherche d'information**. École de bibliothéconomie et des sciences de l'information (EBSI), Université de Montréal, Québec, 2005. *Online*. Disponível em: <http://www.ebsi.umontreal.ca/jetrouve/projet/index.htm>. Acesso em: 02 jan. 2018.

HEIDORN, P. Bryan. The Emerging Role of Libraries in Data Curation and E-science. **Journal of Library Administration**, London, v. 51, n. 7-8, p. 662-672. 2011. <http://dx.doi.org/10.1080/01930826.2011.601269>

HELBIG, Kerstin. Research Data Management Training for Geographers: First Impressions. **ISPRS International Journal of Geo-Information**, Switzerland, v.5, n.4, p. 1–9, 2016. Disponível em: <https://doi.org/10.3390/ijgi5040040>. Acesso em: 02 jan. 2018.

HERNÁNDEZ SAMPIERI, Roberto; FERNÁNDEZ COLLADO, Carlos; BAPTISTA LUCIO, Pilar. **Metodologia de pesquisa**. Porto Alegre: Penso, 2013. 624 p.

HEY, Tony; TANSLEY, Stewart; TOLLE, Kristin. Jim Gray on e-Science: a transformed scientific method. *In*: HEY, Tony; TANSLEY, Stewart; TOLLE, Kristin (Org.). **The Fourth Paradigm: Data-Intensive Scientific Discovery**, 2009. p. xvii - xxxi. Disponível em: <http://itre.cis.upenn.edu/myl/JimGrayOnE-Science.pdf>. Acesso em: 19 nov. 2017.

HEY, Tony, TREFETHEN, Anne. The data deluge: an e-science perspective. *In*: BERMAN, Fran; FOX, Geoffrey; HEY, Tony. **Grid computing: making the global infrastructure a reality**. UK: John Wiley & Sons Chichester, 2003. p.809–824. Doi: 10.1002/0470867167.ch36.

HUGHES, Thomas P. The evolution of larg technological systems. *In*: PINCH, Trevor J. ; BIJKER, Wiebe E.; HUGHES, Thomas P. **The social Construction of Technological Systems: new directions in the sociology and History of Tecnology**.Cambridge (Mass.), MIT, 1993, p. 45-76.

INDUSTRY CANADA. **Seizing Canada’s Moment: Moving Forward in Science, Technology and Innovation**. 2014. 68 p. Disponível em: [https://www.ic.gc.ca/eic/site/113.nsf/vwapj/seizing_moment_ST-I_summary-eng.pdf/\\$file/seizing_moment_ST-I_summary-eng.pdf](https://www.ic.gc.ca/eic/site/113.nsf/vwapj/seizing_moment_ST-I_summary-eng.pdf/$file/seizing_moment_ST-I_summary-eng.pdf). Acesso em: 25 fev. 2018.

INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS. **IEEE International Conference on eScience**. 2014. Disponível em: <https://goo.gl/YUXmQG>. Acesso em: 02 jan. 2018.

INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA (Ibict). **Diretório de Políticas de Acesso Aberto das Revistas Científicas Brasileiras**. 2018. Disponível em: <http://diadorim.ibict.br>. Acesso em: 23 jan. 2018.

INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA (Ibict). **Edital de chamada FINEP/PCAL/XBDB N° 002/2009**. Brasília, 2009. Disponível em: http://www.ibict.br/anexos_noticias /EDITAL_distrib_epto_pcal_xbdb_2_2009.doc. Acesso em: 15 mar. 2018

INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA (Ibict). **Manifesto Brasileiro De Apoio Ao Acesso Livre À Informação Científica**. Brasília: Ibict, 2005. 1 p. Disponível em: <http://livroaberto.ibict.br/Manifesto.pdf>. Acesso em: 30 jun. 2017.

INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA (Ibict). **Manifesto de Acesso Aberto a Dados da Pesquisa Brasileira para Ciência Cidadã.**

Brasília: Ibict, 2016. Disponível em: <http://www.ibict.br/sala-de-imprensa/noticias/item/478-ibict-lanca-manifesto-de-acesso-aberto-a-dados-da-pesquisa-brasileira-para-ciencia-cidada>. Acesso em: 10 fev. 2017.

INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA (Ibict).

Sobre o Ibict: Histórico. 2012. Disponível em: <http://www.ibict.br/sobre-o-ibict/historico>. Acesso em: 08 ago. 2017.

INTER-UNIVERSITY CONSORTIUM FOR POLITICAL AND SOCIAL RESEARCH (ICPSR). **About ICPSR.** Ann Arbor, MI, 2019. [online]. Disponível em:

<http://www.icpsr.umich.edu/icpsrweb/content/about>. Acesso em: 25 abr. 2017.

INTER-UNIVERSITY CONSORTIUM FOR POLITICAL AND SOCIAL RESEARCH (ICPSR). **Guide to Social Science Data Preparation and Archiving:** Best Practice

Throughout the Data Life Cycle. 5. ed. Ann Arbor, MI, 2012. Disponível em: <http://www.icpsr.umich.edu/files/ICPSR/access/dataprep.pdf>. Acesso em: 25 abr. 2017.

JAHNKE, Lori; ASHER, Andrew; KERALIS, Spencer D. C. **The Problem of Data:** Data Management and Curation Practices among University Researchers, Council on Library and Information Resources, Washington, DC, 2012. Disponível em: <https://goo.gl/BPYncb>. Acesso em 20 out. 2017.

JEFFRYES, J. N.; JOHNSTON, L. An E-Learning Approach to Data Information Literacy Education. *In: 120th ASEE Annual Conference & Exposition, Atlanta, Georgia, Anais...* 2013. <https://peer.asee.org/19170>

JISC SERVICES LIMITED. **Meeting the Research Data Challenge.** 2011. 2 p. Disponível em: <https://goo.gl/iCcbuo>. Acesso em 20 set. 2017.

JOHNSON, Andrew M.; BRESNAHAN, Megan M. DataDay! Designing and Assessing a Research Data Workshop for Subject Librarians. **Journal of Librarianship & Scholarly Communication**, Forest Grove, v. 3, n. 2, p. 1–19, 2015. Disponível em: <https://doi.org/10.7710/2162-3309.1229>. Acesso em: 02 jan. 2018.

JOHNSON, Clay A. **The Information Diet:** A Case for Conscious Consumption. Sebastopol, CA: O'Reilly Media. 2012.

JOHNSTON, Lisa R., JEFFRYES, Jon. **Data Management Course for Structural Engineers.** University of Minnesota Libraries: Minnesota, 2012. Disponível em:

https://sites.google.com/a/umn.edu/data-management-course_structures/home-1. Acesso em: 02 jan. 2018.

JORDAN, Michael P. Expanding the invisible college. *In: Annual Meeting Los Angeles*, 36, 1973. Los Angeles. **Anais...**, Los Angeles, 1973.

JUBB, Michael. Introduction: Scholarly communications – disruptions in a complex ecology. 24 p. *In: JUBB, Michael; SHORLEY, Debbie. The Future of Scholarly Communication.* London: Facet, 2013. p. VIII- XXXVI.

KAFEL, Donna; CREAMER, Andrew T.; MARTIN, Elaine R. Building the New England Collaborative Data Management Curriculum. **Journal of eScience Librarianship**, Boston, v. 3, n. 1, p. 60–66. 2014. <https://doi.org/10.7191/jeslib.2014.1066>

KARDELIENE, Laimute; KARDELIS, Kęstutis; BAKUTYTE, Rima. 12. Academic Library as the Space for the Development of Future Specialists' Competence. **Discourse & Communication for Sustainable Education**, v. 5, n. 1, p. 22–30. <https://doi.org/10.2478/dcse-2014-0002.%202014>

KELLAM, Lynda; THOMPSON, Kristi (Eds). **Datalibrarianship: The Academic Data Librarian in Theory and Practice**. Association of College and Research Libraries. Chicago: American Library Association, 2016. 378 p.

KILEY, Robert. The view of the research funder. *In*: JUBB, Michael; SHORLEY, Debbie. **The Future of Scholarly Communication**. London: Facet, 2013. p. 131-144.

KIM, Jeonghyun. Data sharing and its implications for academic libraries. **New Library World**, v. 114, n. 11/12, p. 494-507, out. 2013. <http://dx.doi.org/10.1108/NLW-06-2013-0051>.

KIM, Youngseek. **Institutional and Individual Influences on Scientists' Data Sharing Behaviors**. 2013. 322 fl. Thesis (Doctor of Philosophy) School of Information Studies, Syracuse University, Syracuse, New York. Disponível em: https://surface.syr.edu/cgi/viewcontent.cgi?referer=https://www.google.com/&httpsredir=1&article=1084&context=it_etd. Acesso em: 20 fev. 2019.

KIM, Youngseek. **Scientific Data Reuse Survey [Dataset]**. Inter-university Consortium for Political and Social Research [distributor]. 2017. Disponível em: <http://doi.org/10.3886/E100404V1>. Acesso em: 25 abr. 2017.

KIM, Youngseek. **Scientists' Data Sharing Behaviors [Dataset]**. Inter-university Consortium for Political and Social Research [distributor]. 2016. Disponível em: <http://doi.org/10.3886/E100087V7>. Acesso em: 25 abr. 2017.

KIM, Youngseek; ADDOM, Benjamin K.; STANTON, Jeffrey. M. **Education for eScience Professionals: Integrating Data Curation and Cyberinfrastructure**. United Kingdom, Europe: University of Edinburgh 2011. DOI: <https://doi.org/10.2218/ijdc.v6i1.177>.

KIM, Youngseek; ADLER, Melissa. Social scientists' data sharing behaviors: Investigating the roles of individual motivations, institutional pressures, and data repositories. **International Journal of Information Management**, [s.l.], v. 35, p. 408-418, ago. 2015. Elsevier BV. <https://doi.org/10.1016/j.ijinfomgt.2015.04.007>

KIM, Youngseek, YOON, Ayoung. Scientists' data reuse behaviors: A multilevel analysis. **Journal of the Association for Information Science and Technology**, v.68, n.12, p. 2709-2719. 2017a.

KIM, Youngseek, YOON, Ayoung. Social scientists' data reuse behaviors: Exploring the roles of attitudinal beliefs, attitudes, norms, and data repositories. **Library and Information**

Science Research, v. 39, n. 3, p. 224-233. 2017b.
<http://dx.doi.org/10.1016/j.lisr.2017.07.008>.

KIM, Youngseek; STANTON, Jeffrey. M. Institutional and individual factors affecting scientists' data-sharing behaviors: A multilevel analysis. **Journal of the Association for Information Science & Technology**, Hoboken, New Jersey, v. 67, n. 4, p. 776-799, 2016. DOI: 10.1002/asi.23424.

KIMBROUGH, Julie. L.; GASAWAY, Laura N. Publication of Government-Funded Research, Open Access, and the Public Interest. **Vanderbilt Journal of Entertainment and Technology Law**, Nashville, v. 18, n. 2, p. 267-302, jan. 2016. Disponível em: http://www.jetlaw.org/wp-content/uploads/2016/03/KimbroughGasaway_SPE_7-FINAL.pdf. Acesso em: 06 jan. 2018.

KLING, Rob. The internet and unrefereed scholarly publishing. **Annual Review of Information Science and Technology**, Silver Spring, v. 38, n. 1, jan. 2004, p. 591-631. Disponível em: <https://doi.org/10.1002/aris.1440380113>. Acesso em: 02 jan. 2018.

KOLTAY, Tibor. Are you ready? Tasks and roles for academic libraries in supporting Research 2.0. **New Library World**, [s.l.], v. 117, n. 1-2, p. 94-104, 2016. <https://doi.org/10.1108/NLW-09-2015-0062>

KOLTAY, Tibor. Data literacy: in search of a name and identity. **Journal of Documentation**, [s.l.], v. 71, n. 2, p. 401-415, 2015a. <https://doi.org/10.1108/JD-02-2014-0026>.

KOLTAY, Tibor. Data literacy for researchers and data librarians. **Journal of Librarianship & Information Science**, [s.l.], v. 49, n. 1, p. 3, 2015b. Disponível em: <https://doi.org/10.1177/0961000615616450>. Acesso em: 02 jan. 2018.

KOLTAY, Tibor. Research 2.0 and Research Data Services in academic and research libraries: priority issues. **Library Management**, [s.l.], v. 38, n.6/7, p. 345-353, 2017b. Disponível em: <https://doi.org/10.1108/LM-11-2016-0082>. Acesso em: 02 jan. 2018.

KOLTAY, Tibor. The bright side of information: ways of mitigating information overload. **Journal of Documentation**, [s.l.], v. 73, n. 4, p. 767-775, 2017c. Disponível em: <https://doi.org/10.1108/JD-09-2016-0107>. Acesso em: 02 jan. 2018.

KOLTAY, Tibor. HORSTMANN, Wolfram; WITT, Michael. Data governance, data literacy and the management of data quality. **IFLA Journal**, v.42, n. 4, p. 303-312, 2016. Disponível em: <https://doi.org/10.1177/0340035216672238>. Acesso em: 02 jan. 2018.

KOLTAY, Tibor; SPIRANEC, Sonja; KARVALICS, Laszlo Z. **Research 2.0 and the future of information literacy**. Cambridge: Chandos, 2016. 185 p.

KOLTAY, Tibor; SPIRANEC, Sonja; KARVALICS, Laszlo Z. The Shift of Information Literacy Towards Research 2.0. **Journal of Academic Librarianship**, [s.l.], v. 41, n. 1, p. 87-93, 2015. <http://dx.doi.org/10.1016/j.acalib.2014.11.001>.

KURAMOTO, Hélio. **Currículo Lattes**. 26 set. 2014. Disponível em: <http://goo.gl/4HsZyy>. Acesso em: 25 jul. 2017.

KURAMOTO, Hélio. Informação científica: proposta de um novo modelo para o Brasil. **Ciência da Informação**, Brasília, v. 35, n. 2, 91-102, maio/ago. 2006. Disponível em: <http://goo.gl/aKGVpR>. Acesso em: 10 jul. 2017. <http://dx.doi.org/10.1590/S0100-19652006000200010>.

KURAMOTO, Hélio. **PLS 387/2011 – Projeto de Lei para Acesso Aberto à Produção das Universidades Públicas Brasileiras**. 30 abr. 2013. Disponível em: <http://goo.gl/4mzZHH>. Acesso em: 25 jul. 2017.

KURAMOTO, Hélio. Repositórios Institucionais: políticas e mandatos. *In: SAYÃO et al. Implantação e gestão de repositórios institucionais: políticas, memória, livre acesso e preservação*, 203-217. Salvador: EDUFBA, 2009. Disponível em: <http://goo.gl/jhryVU>. Acesso em: 21 jul. 2017.

LANCASTER, F. Wilfrid. **The measurement and evaluation of the library services**. Washington, D.C.: Information Resources Press, 1977. 395 p.

LATOUR, Bruno. **Ciência em ação: como seguir cientistas e engenheiros sociedade afora**. São Paulo: UNESP, 2000. 438 p.

LATOUR, Bruno. Como terminar uma tese de sociologia. *In: Cadernos de campo*, São Paulo, n.14/15, 2006, p. 339-352. Disponível em: <http://www.bruno-latour.fr/sites/default/files/downloads/90-DIALOGUE-POR.pdf>. Acesso em: 30 set. 2016.

LATOUR, Bruno. Give Me a Laboratory and I will Raise the World. *In: KNORR-CETINA, K. D. ; MULKAY, M. J. Science Observed*. Beverly Hills: Sage, 1983.

LATOUR, Bruno. La Tecnología es la sociedad hecha para que dure. *In: DOMENECH, Miquel; TIRADO, Francisco Javier. Sociologia Simétrica. Ensayos sobre Ciencia, Tecnología y Sociedad*. Barcelona: Gedisa Editorial, 1998, p. 109-142.

LATOUR, Bruno. One more turn after the social turn... *In: MCMULLIN, E. The Social Dimension of Science*. Indiana: Indiana University of Notre Dame Press, 1992, p. 272-294.

LATOUR, Bruno. Os objetos têm história? Encontro de Pasteur com Withehead num banho de ácido láctico. **História, Ciências, Saúde: Manguinhos**, v. II, n.1, p.7-26, mar-jun., 1995.

LATOUR, Bruno. Um coletivo de humanos e não-humanos: No labirinto de Dédalo. *In: LATOUR, Bruno. A Esperança de Pandora: Ensaio Sobre a Realidade dos Estudos Científicos*. Bauru, SP: EDUSC, 2001, p.201-246.

LATOUR, Bruno. **We have never been modern**. Cambridge, MA: Harvard University Press, 1993. 165 p. Disponível em: https://monoskop.org/images/e/e4/Latour_Bruno_We_Have_Never_Been_Modern.pdf. Acesso em: 08 mar. 2017.

LATOURE, Bruno; CALLON, Michel. Unscrewing the big Leviathan: how actors macro-structure reality and how sociologists help them to do so. *In: KNORR-CETINA, K. D.; CICOUREL, A.V. (eds.). **Advances in social theory and methodology: toward and integration of micro- and macro-Sociologies.*** Boston: Routledge; Kegan Paul, 1981, p.277-303. Disponível em: <http://www.bruno-latour.fr/sites/default/files/09-LEVIATHAN-GB.pdf>. Acesso em: 08 mar. 2017.

LAW, John. Technology and Heterogeneous Engineering: The Case of Portuguese Expansion. *In: PINCH, Trevor J. ; BIJKER, Wiebe E.; HUGHES, Thomas P. **The social Construction of Technological Systems: new directions in the sociology and History of Technology.*** Cambridge (Mass.), MIT, 1993, p. 105-127.

LEIDEN UNIVERSITY; ELSEVIER. **Open data: the researcher perspective.** Abr. 2017. 48 p. Disponível em: <https://www.cwts.nl/download/f-53w2.pdf>. Acesso em: 08 ago. 2017.

LEONELLI, Sabina. Why the current insistence on open access to scientific data? Big data, knowledge production, and the political economy of contemporary biology. **Bull Sci Technol Soc**, [s.l.], v. 1-2, n.33, 2013, p. 6–11. doi:10.1177/0270467613496768

LINSINGEN, Irlan von; BAZZO, Walter A.; PEREIRA, Luiz T. V. O que é ciência, tecnologia e sociedade? *In: LINSINGEN, Irlan von; BAZZO, Walter A.; PEREIRA, Luiz T. V. **Introdução aos estudos CTS: ciência, tecnologia e sociedade.*** Espanha: OEI, 2003. p. 119-156 (Cadernos de Ibero-América).

LÓSCIO, Bernadette Farias; BURLE, Caroline; CALEGARI, Newton. **Data on the Web best practices.** W3C Data on the Web Best Practices Working Group, World Wide Web Consortium (W3C), 31 jan. 2017. Disponível em: <https://www.w3.org/TR/dwbp/>. Acesso em: 30 jun. 2018.

LUNA, Sérgio Vasconcelos de. **Planejamento de Pesquisa: uma Introdução.** São Paulo: EDUC, 2000. Disponível em: <http://franciscoqueiroz.com.br/portal/phocadownload/MetodologiadaPesquisa/luna%20sv%20planejamento%20de%20pesquisa.pdf> Acesso em: 15 jan. 2018.

MACHADO, Jorge. Dados abertos e ciência aberta. *In: ALBAGLI, Sarita; MACIEL, Maria Lucia; ABDO, Alexandre Hannud. **Ciência aberta, questões abertas.*** Brasília: IBICT; Rio de Janeiro: UNIRIO, 2015. p. 201-228.

MACY, Katharine. V.; COATES, Heather L. Data information literacy instruction in Business and Public Health. **IFLA Journal**, [s.l.],v. 42, n. 4, p. 313, dez. 2016. DOI: 10.1177/0340035216673382.

MANDINACH, Ellen B.; PARTON, Brennan M.; GUMMER, Edith S.; ANDERSON, Rachel. Ethical and appropriate data use requires data literacy: student data can be a powerful, transformative tool in teaching, but to reap those potential benefits practitioners must become more data literate. **Phi Delta Kappan**, [s.l.], v. 96, p. 25-28, 2015. <https://doi.org/10.1177/0031721715569465>.

MANDINACH, Ellen B.; GUMMER, Edith S. A Systemic View of Implementing Data Literacy in Educator Preparation. **Educational Research**, [s.l.], v. 42, n. 1, p. 30-37, 2013. DOI: 10.3102/0013189X12459803.

MANDINACH, Ellen B.; GUMMER, Edith S. Building a Conceptual Framework for Data Literacy. **Teachers College Record**, New York, v. 117, n. 4, p. 22, abr. 2015.

MARCONI, Marina de Andrade; LAKATOS, Eva Maria. **Metodologia do Trabalho Científico**. 7. ed. São Paulo: Atlas, 2007.

MARCONI, Marina de Andrade; LAKATOS, Eva Maria. **Técnicas de Pesquisa: planejamento e execução de pesquisas; amostragens e técnicas de pesquisa; elaboração, análise e interpretação de dados**. 7 ed. São Paulo: Editora Atlas, 2012. 277 p.

MARQUES, David; ELSEVIER: Library Connect. Research data driving new services. 25 fev. 2013 Disponível em: <https://libraryconnect.elsevier.com/articles/research-data-driving-new-services>. Acesso em: 06 jan. 2018.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY. **Data Management: Workshops**. 2014. Disponível em: <http://libraries.mit.edu/data-management/services/workshops>. Acesso em: 02 jan. 2018.

MATTAR, João. Interações em Ambientes Virtuais de Aprendizagem: histórico e modelos. **Teccogs**, São Paulo, n. 9, p. 53-71, jan.-jun. 2014.

MAUTHNER, Natasha. Are Research Data a Common Resource? **Feminists@ law**, n., v.2,2012. Disponível em: <https://journals.kent.ac.uk/index.php/feministsatlaw/article/view/60/174>. Acesso em: 02 jul. 2018.

MAYBEE, Clarence; CARLSON, Jake; SLEBODNIKC, Maribeth; CHAPMAN, Bert. "It's in the Syllabus": Identifying Information Literacy and Data Information Literacy Opportunities Using a Grounded Theory Approach. **Journal of Academic Librarianship**, [s.l.], v. 41, n. 4, p. 369-376, Jul. 2015. <http://dx.doi.org/10.1016/j.acalib.2015.05.009>.

MAYBEE, Clarence; ZILINSKI, Lisa. Data Informed Learning: A next Phase Data Literacy Framework for Higher Education. *In: 52 Proceedings of the Association for Information Science and Technology*, v.52, n° 1, p. 1-4. **Anais...2015**. Disponível em: <https://doi.org/10.1002/pra2.2015.1450520100108>. Acesso em: 02 jan. 2018.

MCGRATH, Mike. The changing role of the journal editor. p. 117-130. *In: JUBB, Michael; SHORLEY, Debbie. The Future of Scholarly Communication*. London: Facet, 2013. 224 p.

MEADOWS, Arthur Jack. **Communication in science**. London: Butterworth, 1974. 248 p.

MEADOWS, Arthur Jack. **Communicating research**. San Diego: Academic Press, 1998. 256 p.

MEDEIROS, Jackson da S.; CAREGNATO, S. E. Compartilhamento de dados e e-Science:

explorando um novo conceito para a comunicação científica. **Liinc em Revista**, v. 8, n. 2, jul./dez. 2012. Disponível em: <https://goo.gl/dqdk4>. Acesso em: 19 out. 2017. <https://doi.org/10.18617/liinc.v8i2.488>.

MEES, Kenneth. The production of scientific knowledge. **Nature**, [s.l.], v. 100, 1918, p. 355–358.

MOONEY, Hailey; COLLIE, W. Aaron; NICHOLSON, Shawn W.; SOSULSKI, Marya R. Collaborative Approaches to Undergraduate Research Training: Information Literacy and Data Management. **Advances in Social Work**, [s.l.], v. 15, n. 2, 2014, p. 368–89.

MONTEIRO, Elizabete Cristina de Souza de Aguiar. **Direitos Autorais Nos Repositórios De Dados Científicos**: análise sobre os Planos de Gerenciamento dos Dados. 115 f. Programa de Pós-Graduação em Ciência da Informação da Faculdade de Filosofia e Ciências, da Universidade Estadual Paulista – UNESP – Campus de Marília. 20 fev. 2017. Disponível em: <https://goo.gl/QcvkQo>. Acesso em: 02 jan. 2018.

NATIONAL ENDOWMENT FOR THE HUMANITIES. **Data Management Plans for NEH Office of Digital Humanities**. 2018. Disponível em: https://www.neh.gov/sites/default/files/2018-06/data_management_plans_2018.pdf. Acesso em: 13 jan. 2017.

NATIONAL INSTITUTIONS OF HEALTH. **NIH Data Sharing Policy and Implementation Guidance**. 2003. Disponível em: <https://goo.gl/suJsEY>. Acesso em: 25 fev. 2017.

NATIONAL INSTITUTES OF HEALTH. **NIH grants policy statement (10/10) - Part II: Terms and conditions of NIH grant awards, Subpart A: General – File 6 of 6**. 15 out. 2010. Disponível em: <https://goo.gl/q2ECLY>. Acesso em: 25 fev. 2017.

NATIONAL SCIENCE BOARD. **Long-lived digital data collections: enabling research and education in the 21st century**. National Science Foundation, Sept. 2005. Disponível em: <http://www.nsf.gov/pubs/2005/nsb0540/nsb0540.pdf>. Acesso em: 01 set. 2017.

NATIONAL SCIENCE FOUNDATION. **Cyberinfrastructure Vision for 21st Century Discovery**. 2007. Disponível em: <https://goo.gl/AEFEJ4>. Acesso em: 10 abr. 2017.

NATIONAL SCIENCE FOUNDATION. **Dissemination and sharing of research results**. 2010. Disponível em: <https://goo.gl/E6nsXV>. Acesso em: 10 fev. 2018.

NATURE PUBLISHING GROUP. **Availability of data, material, and methods**. 2014. *Online*. Disponível em: <https://goo.gl/2pvKiq>. Acesso em: 10 fev. 2018.

NEW ENGLAND COLLABORATIVE DATA MANAGEMENT CURRICULUM - NECDMC. Lamar Soutter Library, University of Massachusetts Medical School. 2012. Disponível em: <http://library.umassmed.edu/necdmc/index>. Acesso em: 02 jan. 2018.

NEW MEDIA CONSORTIUM. **The NMC Horizon Report 2013 Higher Education**. 2013. 44 p. Disponível em: <https://www.nmc.org/system/files/pubs/1359993875/2013-horizon-report-HE.pdf>. Acesso em: 02 jan. 2018.

OAKLIST Database of Publisher Policies on Open Access. 2009. *Online*. Disponível em: <https://goo.gl/Twz23U>. Acesso em: 26 fev. 2018.

OBJECT MANAGEMENT GROUP. **Business Process Model and Notation (BPMN). V. 2.0**. 2011. Disponível em: <https://www.omg.org/spec/BPMN/2.0/>. Acesso em: 5 dez. 2018.

OFFICE OF SCIENCE AND TECHNOLOGY POLICY. **Memorandum for The Heads of Executive Departments And Agencies**. 22 fev. 2013. 6 p. Disponível em: <https://goo.gl/Ak8UwV>. Acesso em: 10 fev. 2018.

ORGANIZATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT. **OECD principles and guidelines for access to research data from public funding**. Paris: Organization for Economic Co-operation and Development, 2007. Disponível em: <http://www.oecd.org/sti/sci-tech/38500813.pdf>. Acesso em: 31 ago. 2017.

ORGANIZATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT. **Making Open Science a Reality**. (OECD Science, Technology and Industry Policy Papers, 25). Paris: OECD Publishing. 2015. 108 p. Disponível em: <http://dx.doi.org/10.1787/5jrs2f963zs1-en>. Acesso em: 21 out. 2016.

OPEN AIRE. Open Science in Europe. 2015. Disponível em: <https://www.openaire.eu/open-science-europe-overview>. Acesso em: 10 Jan. 2017.

OPEN RESEARCH FUNDERS GROUP. **Prominent funding organizations team up to launch Open Research Funders Group**. Washington, DC, 15 Dez. 2016. Disponível em: <http://www.orfg.org/news/>. Acesso em: 10 Jan. 2017.

OPEN SCIENCE AND RESEARCH INITIATIVE. **The Open Science and Research Handbook**. 2014. 17 p. Disponível em: <https://www.fosteropenscience.eu/sites/default/files/pdf/3986.pdf>. Acesso em: 5 set. 2017.

ORTEGA Y GASSET, José. **Misión del bibliotecario**. Ed. com. Madrid: FUNDACIÓN JOSÉ ORTEGA Y GASSET, 2005. 53p. Disponível em: https://emastromatteo.files.wordpress.com/2010/08/mision_bibliotecario.pdf. Acesso em: 5 set. 2017.

PAGANI, Regina Negri; KOVALESKI, João Luiz; RESENDE, Luis Mauricio. Tics na composição da methodiordinatio: construção de portfólio bibliográfico sobre modelos de transferência de tecnologia. **Ciência da Informação**, Brasília, v. 46, n. 2, 2017. DOI: 10.18225/ci.inf.v47i1.1886. Acesso em: 15 ago. 2019.

PAUWELS C, CLARYSSE B, WRIGHT M, VAN HOVE J. Understanding a new generation incubation model : the accelerator. **Technovation**. Amsterdam: Elsevier Science Bv; n. 50-51, p. 13–24, 2016

PEPE, Alberto; MATTHEW, Mayernik, BORGMAN, Christine L.; SOMPEL, Herbert Van de. From Artifacts to Aggregations: Modeling Scientific Life Cycles on the Semantic Web. **Journal of the American Society for Information Science and Technology**, [s.l.], v.3, n.

61, 2009, p. 567–582. DOI: 10.1002/asi.21263. Disponível em: <https://onlinelibrary.wiley.com/doi/epdf/10.1002/asi.21263>. Acesso em: 5 set. 2017.

PEREIRA, Alice Theresinha Cybis; SCHMITT, Valdenise; DIAS, Regina Álvares C. Ambientes Virtuais de Aprendizagem. In: PEREIRA, Alice Theresinha Cybis (Org.). **Ambientes Virtuais de Aprendizagem**: em diferentes contextos. Rio de Janeiro. Ciência Moderna, 2007. p. 1-22.

PESTALOZZI, Johann Heinrich. **Mes recherches sur la marche de la nature dans l'évolution du genre humain** (1797), présentation, traduction et commentaire par Michel Soëtard — Lausanne : Éditions Payot, 1994. 293 p. Index des notions, (Études et documents).

PIWOWAR, Heather A.; VISION, Todd J. Data reuse and the open data citation advantage. **PeerJ**, v. 1, p. e175, 2013. DOI: <https://doi.org/10.7717/peerj.175>. Disponível em: <https://peerj.com/articles/175.pdf>. Acesso em: 26 jun. 2018.

POOLE, Alex H. The conceptual landscape of digital curation. **Journal of Documentation**, [s.l.], v. 72, n. 5, p. 961, set. 2016. doi: 10.1108/JD-10-2015-0123. Disponível em: <http://dx.doi.org/10.1108/JD-10-2015-0123>. Acesso em: 26 jun. 2017.

PONTIKA, Nancy; KNOTH, Petr; CANCELLIERI, Matteo; PEARCE, Samuel. Fostering Open Science to Research using a Taxonomy and an eLearning Portal. In: **iKnow**: 15th International Conference on Knowledge Technologies and Data Driven Business, 21 - 22 October 2015, Graz, Austria. 2015. Disponível em: http://oro.open.ac.uk/44719/2/kmi_foster_iknow.pdf/. Acesso em: 21 fev. 2018.

PORTELA, Graça. **Pesquisadores aprofundam discussão do acesso livre à informação científica**. 15 mar. 2013. Disponível em: <https://fapergs.rs.gov.br/pesquisadores-aprofundam-discussao-do-acesso-livre-a-informacao-cientifica>. Acesso em: 25 jul. 2017.

PLOS ONE. **Materials and Software Sharing**. [s.d.]. Disponível em: <https://goo.gl/YJ3aJj>. Acesso em: 10 fev. 2018.

PRAHALAD, Coimbatore Krishnarao; HAMEL, Gary. The core competence of the corporation. **Harvard Business Review**, v. 68, n. 3, p. 79-91, maio/jun. 1990.

PROSSER, David C. Researchers and scholarly communications: an evolving interdependency. p. 39-50. In: JUBB, Michael; SHORLEY, Debbie. **The Future of Scholarly Communication**. London: Facet, 2013. 224 p.

PUBLIC LIBRARY OF SCIENCE. **Submission Guidelines**. [s.d.]. Disponível em: <https://goo.gl/Sr1Wof>. Acesso em: 11 set. 2017.

QIN, Jian; D'IGNAZIO, John. Lessons learned from a two-year experience in science data literacy education. In: 31st annual IATUL conference. **Anais...** 2010. Disponível em: <https://goo.gl/jpiJqN>. Acesso em: 19 jan. 2017.

REDE NACIONAL DE ENSINO E PESQUISA. **Carta-Convite 01/2017**. 15 de março de 2017. 2017a. Disponível em: <https://goo.gl/QGyjL1>. Acesso em: 11 ago. 2017a.

REDE NACIONAL DE ENSINO E PESQUISA. GT Dados Abertos (GT-AADP / GT-RDP): **Sobre o Projeto**. 2017b. Disponível em: https://wiki.rnp.br/x/JS_WBQ. Acesso em: 08 ago. 2017.

REIS, Edna Afonso; REIS, Ilka Afonso. **Análise Descritiva de Dados: Tabelas e Gráficos**. Relatório Técnico RTE-04/2001, Universidade Federal de Minas Gerais, Instituto de Ciências Exatas Departamento de Estatística: Belo Horizonte, 2001. *Online*. Disponível em: <http://www.leg.ufpr.br/~silvia/CE055/rte0104.pdf>. Acesso em: 27 nov 2019.

RESEARCH DATA ALLIANCE. **RDA in a Nutshell**: August 2017. 2017. Disponível em: <https://www.rd-alliance.org/node/51727>. Acesso em: 11 ago. 2017.

RESEARCH DATA ALLIANCE. **Who is RDA?** Disponível em: <https://www.rd-alliance.org/node/51727>. Acesso em 03 dez. 2017.

RESEARCH DATA CANADA. **Research data management in Canadian universities: a statement of principles**. 2016. 5p. Disponível em: <https://goo.gl/f6cRaR>. Acesso em: 12 jan. 2018.

RESEARCH DATA NETHERLANDS. **Essentials 4 Data Support**, [s.d.], *Online*. Disponível em: <http://datasupport.researchdata.nl/en/>. Acesso em: 20 fev. 2019.

RUSBRIDGE, Chris. **Representation Information: what is it and why is it important?**. 2007. Disponível em: <http://www.dcc.ac.uk/news/representation-information-what-it-and-why-it-important>. Acesso em: 31 ago. 2016.

SURKIS, Alisa; LAPOLLA, Fred Willie Zametkin; CONTAXIS, Nicole; READ, Kevin B. 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**, Pittsburgh, v.105, n. 2, 2017, p 185–191. DOI: [dx.doi.org/10.5195/jmla.2017.35](https://doi.org/10.5195/jmla.2017.35).

REGISTRY OF OPEN ACCESS REPOSITORIES. 2018. Acesso em 03 fev. 2018. Disponível em: <http://roar.eprints.org/content.html>. Acesso em: 12 jan. 2018.

RITTEL, Horst. W; WEBBER, Melvin. M. Dilemmas in a general theory of planning. **Policy Sciences**, *online*, v. 4 n. 2, p. 155-169. 1973. Disponível em: <https://goo.gl/uZSSqL>. Acesso em: 12 dez. 2017.

ROARMAP. Registry of Open Access Repository Material Archiving Policies. 2018. Disponível em: <https://goo.gl/260Bge>. Acesso em: 12 jan. 2018.

ROCHA, Rafael Port da. **Proposta para Grupo de Trabalho sobre Acesso Aberto a Dados de Pesquisa RDP Brasil - Rede de Dados de Pesquisa Brasileira**. 2017. Disponível em: <https://goo.gl/B3bd7F>. Acesso em 3 ago. 2017.

RODRIGUES, Eloi. As universidades e o open access: apresentação para dirigentes universitários. *In*: CONFERÊNCIA OPEN ACCESS, 3, 2008 Braga. **Apresentação...** Braga: Universidade do Minho: 15 a 16 dez. 2008. Disponível em: <http://confoa08.sdum.uminho.pt/apresentacoes/Apresentação%20RCAAP2.pdf>. Acesso em: 15 jul. 2017.

ROOSENDAAL, Hans E.; GEURTS, Petrus A.T.M. Forces and functions in scientific communication: an analysis of their interplay. *In: CRISP 97 Cooperative Research Information Systems in Physics, University of Oldenburg, Germany, 1997.* p. 1-32. Disponível em: <https://goo.gl/BQLNik>. Acesso em: 27 jan. 2018.

ROYAL NETHERLANDS ACADEMY OF ARTS AND SCIENCES (KNAW). **Replication studies: improving reproducibility in the empirical sciences.** Amsterdam, KNAW. 2017. 68 p. Disponível em: <https://goo.gl/j8R1kb>. Acesso em 3 abr. 2018.

ROYAL SOCIETY. Policy Studies Unit. **Science as an Open Enterprise**, Great Britain, 2012. 105 p. Disponível em: <https://goo.gl/49ZUMH>. Acesso em 3 ago. 2017.

SAYÃO, Luis; MARCONDES, Carlos Henrique. À guisa de introdução: repositórios institucionais e livre acesso. *In: SAYÃO, Luis; TOUTAIN, Lídia Brandão; ROSA, Flavia Garcia; MARCONDES, Carlos Henrique. **Implantação e gestão de repositórios institucionais: políticas, memória, livre acesso e preservação**, 203-217. Salvador: EDUFBA, 2009.* Disponível em: <http://goo.gl/jhryVU>. Acesso em: 21 jul. 2017.

SAYÃO, Luis Fernando; SALES, Luana Farias. Algumas considerações sobre os Repositórios digitais de dados de Pesquisa. **Inf. Inf.**, Londrina, v. 21, n. 2, p. 90-115, maio/ago., 2016.

SAYÃO, Luis Fernando; SALES, Luana Farias. Ciberinfraestrutura de informação para a pesquisa: uma proposta de arquitetura para integração de repositórios e sistemas CRIS. **Inf. & Soc.:Est.**, João Pessoa, v.25, n.3, p. 163-184, set./dez. 2015.

SAYÃO, Luis Fernando; SALES, Luana Farias. Curadoria digital: um novo patamar para preservação de dados digitais de pesquisa. **Inf. & Soc.:Est.**, João Pessoa, v.22, n.3, p. 179-191, set./dez. 2012.

SAYÃO, Luis Fernando; SALES, Luana Farias. Dados abertos de pesquisa: ampliando o conceito de acesso livre. **RECIIS – Rev. Eletron. de Comun. Inf. Inov. Saúde**, v.8, n.2, p.76-92. jun. 2014. Disponível em: DOI:10.3395/reciis.v8i2.934.pt. Acesso em 3 set. 2017.

SAYÃO, Luis Fernando; SALES, Luana Farias. **Guia de Gestão de Dados de Pesquisa para Bibliotecários e Pesquisadores.** Rio de Janeiro: CNEN, 2015. 90 p. Disponível em: <https://goo.gl/6JBkKV>. Acesso em 3 ago. 2017.

SAYÃO, Luis Fernando; SALES, Luana Farias. O impacto da curadoria digital dos dados de pesquisa na comunicação científica. **Encontros Bibli: revista eletrônica de Biblioteconomia e Ciência da Informação**, Florianópolis, v. 17, n. esp. 2 – III SBCC, p.118-135, 2012. DOI:10.5007/1518-2924.2012v17nesp2p118

SCIENCE-METRIX. **Analytical Support for Bibliometrics Indicators: Open access availability of scientific publications.** Jan. 2018. Montréal. 36p. Disponível em: <https://goo.gl/B9DbL9>. Acesso em: 26 fev. 2018.

SCHNEIDER, R. Research data literacy. In KURBANOGLU, S., *et al.* (Eds.), **Worldwide commonalities and challenges in information literacy research and practice**. Cham: Springer International, v. 397, p. 134-140, 2013.

SCHONFELD, Roger C. The library users' view. p. *In*: JUBB, Michael; SHORLEY, Debbie. **The Future of Scholarly Communication**. London: Facet, 2013. p.169-178.

SETENARESKI, Ligia Eliana; SUNYE, Marcos Sfair; SHIMA, Walter. Adopting Open Protocols to Increase the Impact on Digital Repositories. In Universal Access in Human-Computer Interaction. Applications and Services for Quality of Life: 7th International Conference, UAHCI 2013, Held as Part of HCI International 2013, Las Vegas, NV, USA, 21-26 jul., 2013, **Proceedings...**, Part III. Lecture Notes in Computer Science. V. 8011, Springer Berlin Heidelberg, p. 408-416.

SHERPA RoMEO. **Publisher copyright policies & self-archiving**. *Online*. Disponível em: <https://goo.gl/QpGert>. Acesso em: 26 fev. 2018.

SHORISH, Yasmeen. Data information literacy and undergraduates: A critical competency. **College & Undergraduate Libraries**, [s.l.], v. 22, n. 1, p. 97-106, 2015. DOI: 10.1080/10691316.2015.1001246

SISMONDO, Sergio. Actor Network Theory. *In*: SISMONDO, Sergio. **An Introduction to Science and Technology Studies**. Wiley Blackwell: West Sussex, UK, 2010 p.81-92

SOCIAL SCIENCES AND HUMANITIES RESEARCH COUNCIL OF CANADA. **Research Data Archiving Policy**. 1990. Disponível em: <https://goo.gl/Pyk9ZF>. Acesso em: 25 fev. 2018.

TASK GROUP ON DATA CITATION STANDARDS AND PRACTICES, CODATA-ICSTI. Out of Cite, Out of Mind: The Current State of Practice, Policy, and Technology for the Citation of Data. **Data Science Journal**, [s.l.], v. 12, p. CIDCR1-CIDCR75, 2013. Disponível em: <https://goo.gl/ZZGBGi>. Acesso em: 26 jun. 2018.

TAYLOR, John. M. The UK eScience programme [Powerpoint presentation]. *In*: ESCIENCE LONDON MEETING. Set. 2001.

TENOPIR, Carol; ALLARD, Suzie; DOUGLASS, Kimberly; AYDINOGLU, Arsev Umur; WU, Lei; READ, Eleanor; MANOFF, Maribeth; FRAME, Mike. Data Sharing by Scientists: Practices and Perceptions. **PLoS ONE**, [s.l.], v. 6, n. 6, p. 1-21, 2011

TENOPIR, Carol; BIRCH, Ben; ALLARD, Suzie. **Academic libraries and research data services: current practices and plans for futur - An ACRL White paper**. ACRL, 2012. Disponível em: <https://goo.gl/eoQsXH>. Acesso em 20 Out. 2017. 56p.

TENOPIR, Carol; DALTON, Elizabeth D.; ALLARD, Suzie; FRAME, Mike; PJESIVAC, Ivanka; BIRCH, Ben; POLLOCK, Danielle; DORSETT, Kristina. Changes in Data Sharing and Data Reuse Practices and Perceptions among Scientists Worldwide. **PLoS ONE**, [s.l.], v.10, n8, p.1-24, 2015a. Disponível em: <https://doi.org/10.1371/journal.pone.0134826>. Acesso em: 02 jan. 2018.

TENOPIR, Carol, HUGHES, Dane; ALLARD, Suzie; FRAME, Mike; BIRCH, Ben; BAIRD, Lynn; SANDUSKY, Robert J; LANGSETH, Madison; LUNDEEN, Andrew. Research data services in academic libraries: Data intensive roles for the future? **Journal of eScience Librarianship**, Boston, v. 4, n. 2, 2015b, 23 p. Disponível em: <http://dx.doi.org/10.7191/jeslib.2015.1085>. Acesso em: 20 out. 2017.

TENOPIR, Carol; KING, Donald W. The Growth of Journal Publishing. *In*: COPE, Bill; PHILIPS, Angus. (eds) **The Future of the Academic Journal**. 2 ed. Oxford: Chandos. 2014. p. 159-177.
http://neamathisi.com/_uploads/Cope__Kalantzis_Changing_Knowledge_Ecologies_2014.pdf

TENOPIR, Carol; POLLOCK, Danielle; ALLARD; Suzie; HUGHES, Dane. Research data services in European and North American libraries: Current offerings and plans for the future. **Proceedings of the Association for Information Science & Technology** v.53, n.1, p. 1. 2016.

TENOPIR, Carol, SANDUSKY, Robert J., ALLARD; Suzie; BIRCH, Ben. Academic librarians and research data services: preparation and attitudes. **International Federation of Library Associations and Institutions**, v. 39, n. 1, p. 70–78, 2013. DOI: 10.1177/0340035212473089.

TENOPIR, Carol; SANDUSKY, Robert J; ALLARD, Suzie; BIRCH, Ben. Research data management services in academic research libraries and perceptions of librarians. **Library and Information Science Research**, [s.l.], n. 36, v. 2, 84, 2014.
<https://doi.org/10.1016/j.lisr.2014.11.003>

TENOPIR, Carol; TALJA, Sanna; HORSTMANN, Wolfram; LATE, Elina; HUGHES, Dane; POLLOCK, Danielle; SCHMIDT, Birgit; BAIRD, Lynn ; SANDUSKY, Robert J.; ALLARD, Suzie. 1. Research Data Services in European Academic Research Libraries. **Liber Quarterly: The Journal of European Research Libraries**,[s.l.], v. 27, n. 1, p. 23-44, 2017.

THE ENGINEERING AND PHYSICAL SCIENCES RESEARCH COUNCIL. **EPSRC policy framework on research data**. Disponível em: <https://goo.gl/8jDswQ>>. Acesso em: 22 fev. 2018.

THE WELCOME TRUST. **Open access policy**. 2007. Disponível em: <https://wellcome.ac.uk/node/1925>. Acesso em: 25 fev. 2017.

THE WELCOME TRUST. **Policy on data, software and materials management and sharing**. 2017. Disponível em: <https://wellcome.ac.uk/node/1912>. Acesso em: 25 fev. 2017.

TRI-AGENCY Statement of Principles on Digital Data Management. Canadian Institutes of Health Research (CIHR). The Natural Sciences and Engineering Research Council of Canada (NSERC). The Social Sciences and Humanities Research Council of Canada (SSHRC): Alberta, 2014. 5 p. Disponível em: <https://goo.gl/ZuTZ6J>. Acesso em: 25 fev. 2018.

UHLIR, Paul; SCHRÖDER, Peter. Open Data for Global Science. **Data Science Journal**, [s.l.], Jun. 2007; v. 6 (Open Data Issue). Disponível em: <https://doi.org/10.2481/dsj.6.p.OD36-OD56>. Acesso em: 01 dez. 2017.

UNITED NATIONS EDUCATIONAL, SCIENTIFIC AND CULTURAL ORGANISATION (UNESCO). Science Report: Towards 2030. Bruxelas. Nov. 2015. 820 p. Disponível em: <https://goo.gl/VnQtLb>. Acesso em: 27 jul. 2018.

UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ. **Programa de Pós-Graduação em Tecnologia e Sociedade – PPGTE**: Apresentação. *Online*. 2018. Disponível em: <https://goo.gl/SzyEMG>. Acesso em: 04 fev. 2018.

UNIVERSITY OF BATH. Terminology for Research Data Management. 2011. Disponível em: <https://goo.gl/6tjQcw>. Acesso em: 01 dez. 2017.

UNIVERSITY OF EDINBURGH; EDINA Data Library. Research Data Service staff in Information Services. **Research Data and Management Training (MANTRA)** [online course]. Out. 2017. Disponível em: <https://mantra.edina.ac.uk/>. Acesso em: 20 fev. 2019.

UNIVERSITY OF EDINBURGH. **Our definitions**: Research data 2016. Disponível em: <https://goo.gl/wznsyF>. Acesso em: 01 mar. 2017.

UNIVERSITY OF EDINBURGH. **Why research data management?** 2015. *Online*. Disponível em: <https://goo.gl/KsyrT1>. Acesso em: 15 mar. 2017.

UNIVERSITY OF MINNESOTA. Data Management Course: Winter 2014 Workshop Series. 2014. Disponível em: <https://sites.google.com/a/umn.edu/data-management-workshop-series>. Acesso em: 08 mar. 2017.

UNIVERSITY OF OXFORD. Bodleian Data Library. **Workshops and Classes**: Research Data Management handouts. 2018-2019. Disponível em: <https://libguides.bodleian.ox.ac.uk/workshops/rdm>. Acesso em: 20 fev. 2019.

VACCAREZZA, Leonardo Silvio. Ciencia, tecnología y sociedad: el estado de lá cuestión en América Latina. **Revist@ do Observatório do Movimento pela Tecnologia Social da América Latina**, [s.l.], v. 01, n. 01, p. 42-64, jul. 2011.

VARVEL Jr., Virgil E.; BAMMERLIN, Elin J.; PALMER, Carole L. Education for Data Professionals: A Study of Current Courses and Programs. In **Proceedings of the 2012 iConference**, 527–529. Toronto, Ontario, Canada: ACM, 2012. Disponível em: <http://dx.doi.org/10.1145/2132176.2132275>. Acesso em: 02 jan. 2018.

VEIGA, Viviane Santos de Oliveira. **A percepção dos pesquisadores portugueses e brasileiros da área de Neurociências quanto ao compartilhamento de artigos científicos e dados de pesquisa no acesso aberto verde**: custos, benefícios e fatores contextuais - Rio de janeiro, 2017. Tese (Doutorado) – Instituto de Comunicação e Informação Científica e Tecnológica em Saúde, Pós-Graduação em Informação e Comunicação em Saúde, 2017. 293 f.

VELTEROP, Jan. Holy Cow, Peer Review, 2012. Disponível em: <https://goo.gl/gvytMx>. Acesso em: 08 mar. 2017.

VERBAAN, Eddy; COX, Andrew M. Occupational sub-cultures, jurisdictional struggle and Third Space: Theorizing professional service responses to Research Data Management.

Journal of Academic Librarianship, [s.l.], v.40, n. 3/4, p. 211–219, 2014. Disponível em: <https://goo.gl/aZntMR>. Acesso em: 28 fev. 2017. <https://doi.org/10.1016/j.acalib.2014.02.008>.

VITAE. **About the Vitae Researcher Development Framework**. 2011, *online*. Disponível em: <https://goo.gl/gJAzwa>. Acesso em: 03 mar. 2018.

VITAE. **The Vitae Researcher Development Framework**. 2018, *online*. Disponível em: <https://goo.gl/XiXfVb>. Acesso em: 05 jan. 2018.

WALLIS, Jillian C.; ROLANDO, Elizabeth; BORGMAN, Christine L. If We Share Data, Will Anyone Use Them? Data Sharing and Reuse in the Long Tail of Science and Technology. (Research Article). **PLoS ONE**, [s.l.], v. 8, n. 7, p. 17, 2013. doi:10.1371/journal.pone.0067332.

WANNER, Amanda. Data literacy instruction in academic libraries: best practices for librarians. **Archival and Information Studies Student Journal**, [s.l.], The UBC School of Library, n.1, 2015. Disponível em: <http://ojs.library.ubc.ca/index.php/seealso/article/view/186335/185517>. Acesso em: 02 fev. 2017.

WE THE PEOPLE. **Require free access over the Internet to scientific journal articles arising from taxpayer-funded research**. 20 maio 2012. Disponível em: <https://petitions.obamawhitehouse.archives.gov/node/113476>. Acesso em: 10 fev. 2018.

WEB OF SCIENCE - Coleção Principal (Clarivate Analytics). Detalhes da Base – descrição. 2017. Disponível em: <https://goo.gl/krDarD>. Acesso em: 30 nov. 2017.

WHITMIRE, Amanda L. Implementing a Graduate-Level Research Data Management Course: Approach, Outcomes, and Lessons Learned. **Journal of Librarianship & Scholarly Communication**, [s.l.], v.3, n. 2, p. 1–22, 2015. Disponível em: <https://doi.org/10.7710/2162-3309.1246>. Acesso em: 02 jan. 2018.

WITT, Michael. Co-designing, Co-developing, and Co-implementing an Institutional Data Repository Service. **Journal of Library Administration**, [s.l.], v. 52, n. 2, 2012, p. 172–188. DOI: 10.1080/01930826.2012.655607. Disponível em: <https://www.tandfonline.com/doi/full/10.1080/01930826.2012.655607>. Acesso em: 02 jan. 2018.

WITT, Michael; HORSTMANN, Wolfram; CLEMENT, Ryan; BLAU, Amy; ABBASPOUR, Parvaneh; GANDOUR-ROOD, Eli. Team-based data management instruction at small liberal arts colleges. **IFLA Journal**, v. 43, n. 1, p. 105 – 118, 2017. DOI: 10.1177/0340035216678239. Disponível em: <https://journals.sagepub.com/doi/10.1177/0340035216678239>. Acesso em: 02 jan. 2018.

WHYTE, Angus; JONES, Sarah; PRYOR, Graham. **Delivering Research Data Management Services: Fundamentals of Good Practice**. London: Facet Publishing, 2014. 224 p.

WOOD, John Coping with the Data Deluge. *In*: JUBB, Michael; SHORLEY, Debbie. **The Future of Scholarly Communication**. London: Facet, 2013. p. 75-87.

ZOTERO. **Documentation**. 2018. Disponível em: <https://www.zotero.org/support/>. Acesso em: 06 jan. 2018.

APÊNDICE A – Referências análise sistemática – teses e dissertações

Quadro 30 - Teses e dissertações selecionadas – resumo

Autor	Ano	Nível	Instituição	Área De Concentração	Disponível Em:
ALVES, Ana Paula Meneses	2016	Doutorado	UNESP	Ciência da Informação	https://goo.gl/Fxh74q
AMORIM, Iara Rodrigues de	2013	Mestrado	UFRJ	Ciência da Informação	https://goo.gl/2KnWM1
APPEL, Andre Luiz	2014	Mestrado	UFRJ	Ciência da Informação	https://goo.gl/iBkf8C
CARVALHO, Erika Rayanne Silva De.	2018	Mestrado	UNB	Ciência da Informação	https://bit.ly/2k8Nrqq
COELHO, Vânia Lúcia	2017	Mestrado	UFSCar	Ciência, Tecnologia e Sociedade	https://goo.gl/dW38ei
COSTA, Maíra Murrieta	2017	Doutorado	UNB	Ciência da Informação	https://goo.gl/BMSHb4
FIGUEIREDO, Eduarda Ferreira.	2018	Mestrado	UFPR	Ciência da Informação	https://bit.ly/2lABg5X
JORGE, Vanessa De Arruda	2018	Doutorado	UFPR	Ciência da Informação	https://bit.ly/2k8NpPe
MACHADO, Denise Ramires	2015	Mestrado	UFGRS	Comunicação E Informação	https://goo.gl/Bp1ith
MAIA, Louise Pereira.	2018	Mestrado	UFRJ	Ciência da Informação	https://bit.ly/2kxDrHj
MARQUES, Luana Farias Sales	2014	Doutorado	UFRJ	Ciência da Informação	https://goo.gl/9PJyX7
MEDEIROS, Jackson Da Silva	2014	Doutorado	UFGRS	Comunicação E Informação	https://goo.gl/x4PC9C
MONTEIRO, Elizabete Cristina De Souza De Aguiar	2017	Mestrado	UNESP	Ciência da Informação	https://goo.gl/QcvkQo
NOVELLI, Valéria Aparecida Moreira	2012	Mestrado	UFSCar	Ciência, Tecnologia e Sociedade	https://goo.gl/73iG3D
OLIVEIRA, Adriana Silva de	2016	Doutorado	UFPE	Ciência da Informação	https://goo.gl/s9s7bb
PRADO, Alysson Bolognesi	2017	Doutorado	UNICAMP	Ciência da Computação	https://goo.gl/1J1Xgm
SANTOS, Daiane Barrili Dos	2018	Mestrado	UFGRS	Comunicação E Informação	https://bit.ly/2klrMvf
SILVA, Daniel Lins da	2017	Doutorado	USP	Sistemas Digitais	https://goo.gl/KzgaUF
SOUSA, Margarida Maria de	2014	Doutorado	USP	Ciência da Informação	https://goo.gl/zjQL1s
SOUZA, Thales Vicente de	2016	Mestrado	UFF	Ciência da Informação	https://goo.gl/Dzx54r
TYGEL, Alan Freihof	2016	Doutorado	UFRJ	Informática	https://goo.gl/d1YS5a

Fonte: Autoria Própria (2018).

ALVES, Ana Paula Meneses. **Competência Informacional e o uso ético da informação na produção científica: o papel do bibliotecário na produção intelectual no ambiente acadêmico.** 2016. Disponível em: <https://goo.gl/Fxh74q>. Acesso em: 02 jan. 2018.

AMORIM, Iara Rodrigues de. **Intelligence-based information literacy.** 2013. 145 f. Dissertação (Mestrado em Ciência da Informação) - Universidade Federal do Rio de Janeiro / Instituto Brasileiro de Informação em Ciência e Tecnologia, Rio de Janeiro, 2013. Disponível em: <https://goo.gl/2KnWM1>. Acesso em: 02 jan. 2018.

APPEL, Andre Luiz. **A E-Science E As Atuais Práticas De Pesquisa Científica'** 24/03/2014 88 f. Mestrado em CIÊNCIA DA INFORMAÇÃO - UFRJ - IbiCT Instituição de Ensino: UNIVERSIDADE FEDERAL DO RIO DE JANEIRO, Rio de Janeiro Biblioteca Depositária: UFRJ. Disponível em: <https://goo.gl/iBkf8C>. Acesso em: 02 jan. 2018.

CARVALHO, Erika Rayanne Silva De. **Diferenças na produção, compartilhamento e (re)uso de dados: a percepção de pesquisadores da química, antropologia e educação'** 20/02/2018 173 f. Mestrado em Ciências Da Informação Instituição de Ensino: UNIVERSIDADE DE BRASÍLIA, Brasília Biblioteca Depositária: Biblioteca Central da UnB. Disponível em: <https://bit.ly/2k8Nrqq>. Acesso em: 30 Ago. 2019.

COELHO, Vânia Lúcia . **Relação do capital cultural e da competência em informação para o acesso às redes sociais sobre informação científica.** Disponível em: <https://goo.gl/dW38ei>. Acesso em: 02 jan. 2018.

COSTA, Maíra Murrieta. **Diretrizes para uma política de gestão de dados científicos no Brasil.** 2017. 288 f., il. Tese (Doutorado em Ciência da Informação) - Universidade de Brasília, Brasília, 2017. Disponível em: <https://goo.gl/BMSHb4>. Acesso em: 02 jan. 2018.

FIGUEIREDO, Eduarda Ferreira. **Repositórios digitais de publicações ampliadas: Um estudo de caso'** 27/02/2018 69 f. Mestrado em Ciência Da Informação Instituição de Ensino: Universidade Federal De Pernambuco, Recife. Disponível em: <https://bit.ly/2lABg5X>. Acesso em: 30 Ago. 2019.

JORGE, VANESSA DE ARRUDA. **Abertura e compartilhamento de dados para pesquisa nas situações de emergência em saúde pública: o caso do vírus zika'** 12/09/2018 263 f. Doutorado em Ciência Da Informação - Ufrj - IbiCT Instituição De Ensino: Universidade Federal Do Rio De Janeiro, Rio de Janeiro Biblioteca Depositária: <http://ridi.ibict.br/handle/123456789/993>. Disponível em: <https://bit.ly/2k8NpPe>. Acesso em: 30 Ago. 2019.

MACHADO, DENISE RAMIRES. **Dados de pesquisa em repositório institucional: o caso do Edinburgh DataShare,** 12/05/2015. 113 f. Mestrado em Comunicação E Informação Instituição de Ensino: Universidade Federal Do Rio Grande Do Sul, Porto Alegre Biblioteca Depositária: Biblioteca Setorial FABICO. Disponível em: <https://goo.gl/Bp1i1h>. Acesso em: 02 jan. 2018.

MAIA, Louise Pereira. **Curadoria Digital em Bibliotecas: uma aplicação do modelo do DCC no planejamento de criação da Biblioteca MUSAL Digital.'** 28/12/2018 145 f. Mestrado Profissional em biblioteconomia Instituição de Ensino: Universidade Federal Do Estado Do

Rio De Janeiro, Rio de Janeiro. Disponível em: <https://bit.ly/2kxDrHj>. Acesso em: 30 Ago. 2019.

MARQUES, Luana Farias Sales. **Integração Semântica De Publicações Científicas E Dados De Pesquisa:** Proposta De Modelo De Publicação Ampliada Para A Área De Ciências Nucleares. 23/07/2014 264 f. Doutorado em Ciência Da Informação - UFRJ - Ibict Instituição de Ensino: Universidade Federal Do Rio De Janeiro, Rio de Janeiro Biblioteca Depositária: UFRJ. Disponível em: <https://goo.gl/9PJyX7>. Acesso em: 02 jan. 2018.

MEDEIROS, JACKSON DA SILVA. **Uma Investigação Sobre A Autoria De Dados Científicos:** teias de uma rede em construção' 06/05/2015 192 f. Doutorado em Comunicação E Informação Instituição de Ensino: Universidade Federal Do Rio Grande Do Sul, Porto Alegre Biblioteca Depositária: Biblioteca Setorial da Faculdade de Biblioteconomia e Comunicação. Disponível em: <https://goo.gl/x4PC9C>. Acesso em: 02 jan. 2018.

MONTEIRO, ELIZABETE CRISTINA DE SOUZA DE AGUIAR. **Direitos Autorais Nos Repositórios De Dados Científicos:** análise sobre os Planos de Gerenciamento dos Dados. 20/02/2017 115 f. Mestrado em Ciência Da Informação Instituição De Ensino: Universidade Est.Paulista Júlio De Mesquita Filho/Marília, Marília Biblioteca Depositária: Campus de Marília. Disponível em: <https://goo.gl/QcvkQo>. Acesso em: 02 jan. 2018.

NOVELLI, Valéria Aparecida Moreira. Ferramentas aplicáveis à mediação do acesso, busca e aprendizagem do uso de fontes de informação em bibliotecas universitárias. 2012. 165 f. Dissertação (Mestrado em Multidisciplinar) - Universidade Federal de São Carlos, São Carlos, 2012. Disponível em: <https://goo.gl/73iG3D>. Acesso em: 02 jan. 2018.

OLIVEIRA, Adriana Silva de Oliveira. Desvendando a autoralidade colaborativa na e-science sob A ótica dos direitos de propriedade intelectual. 2016. 300 f. Dissertação (Doutorado em Ciência da Informação) - Universidade Federal da Paraíba, João Pessoa, 2016. Disponível em: <https://goo.gl/s9s7bb>. Acesso em: 02 jan. 2018.

PRADO, ALYSSON BOLOGNESI. **Uma Análise do Papel de Sistemas Computacionais como Mediadores em Redes Sociotécnicas Científicas sob a Ótica da Actor-Network Theory'** 07/03/2017 148 f. Doutorado em Ciência Da Computação Instituição De Ensino: Universidade Estadual De Campinas, Campinas Biblioteca Depositária: undefined. Disponível em: <https://goo.gl/1J1Xgm>. Acesso em: 02 jan. 2018.

SANTOS, Daiane Barrili Dos. **Publicações ampliadas:** aspectos da integração de dados de pesquisa' 18/01/2018 187 f. Mestrado em Programa de Pós-Graduação em Comunicação Instituição de Ensino: Universidade Federal Do Rio Grande Do Sul, Porto Alegre Biblioteca Depositária: Biblioteca Setorial da Faculdade de Biblioteconomia e Comunicação. Disponível em: <https://bit.ly/2klrMvf>. Acesso em: 30 Ago. 2019.

SILVA, Daniel Lins da. **Estratégia computacional para apoiar a reprodutibilidade e reúso de dados científicos baseado em metadados de proveniência.** 2017. Tese (Doutorado em Sistemas Digitais) - Escola Politécnica, Universidade de São Paulo, São Paulo, 2017. Disponível em: <http://www.teses.usp.br/teses/disponiveis/3/3141/tde-05092017-095907/>. Acesso em: 2018-01-02. Disponível em: <https://goo.gl/KzgaUF>. Acesso em: 02 jan. 2018.

SOUSA, Margarida Maria de. A função educativa do bibliotecário no século XXI: desafios para sua formação e atuação. 2014. Disponível em: <https://goo.gl/zjQL1s>. Acesso em: 02 jan. 2018.

SOUZA, Thales Vicente De. **Curadoria Digital: um novo espaço de atuação do profissional de informação'** 30/05/2016 102 f. Mestrado em Ciência Da Informação Instituição De Ensino: Universidade Federal Fluminense, Niterói Biblioteca Depositária: Biblioteca Central do Gragoatá – UFF. Disponível em: <
<https://goo.gl/Dzx54r>. Acesso em: 02 jan. 2018.

TYGEL, Alan Freihof. **Semantic Tags for Open Data Portals: Metadata Enhancements for Searchable Open Data'** 21/07/2016 90 f. Doutorado em Informática Instituição De Ensino: Universidade Federal Do Rio De Janeiro, Rio de Janeiro Biblioteca Depositária: NCE/UFRJ. Disponível em: <https://goo.gl/d1YS5a>. Acesso em: 02 jan. 2018.

APÊNDICE B – Autores e artigos mais citados

A distribuição da quantidade de registros recuperados na estratégia de busca por autores, em cada base de dados, pode ser vista na Tabela 17.

Tabela 17 - Pesquisa por autores – registros recuperados

Busca por Autores com maior produção	Referências Capes	Referências EBSCOHost	Referências Web of Science	Total Referências	% de referência por autor	Referências Selecionadas	% referências selecionadas
Borgman	19	17	18	54	18,82%	54	18,82%
Koltay	27	37	12	76	26,48%	76	26,48%
Tenopir	42	85	30	157	54,70%	157	54,70%
Total	88	139	60	287	100,00%	287	100,00%
referências % de referências por bases	30,66%	48,43%	20,91%	100,00%			

Fonte: Autoria própria (2019).

A distribuição da quantidade de registros selecionados na estratégia de busca por autores de acordo com os anos de publicação pode ser vista na Tabela 18.

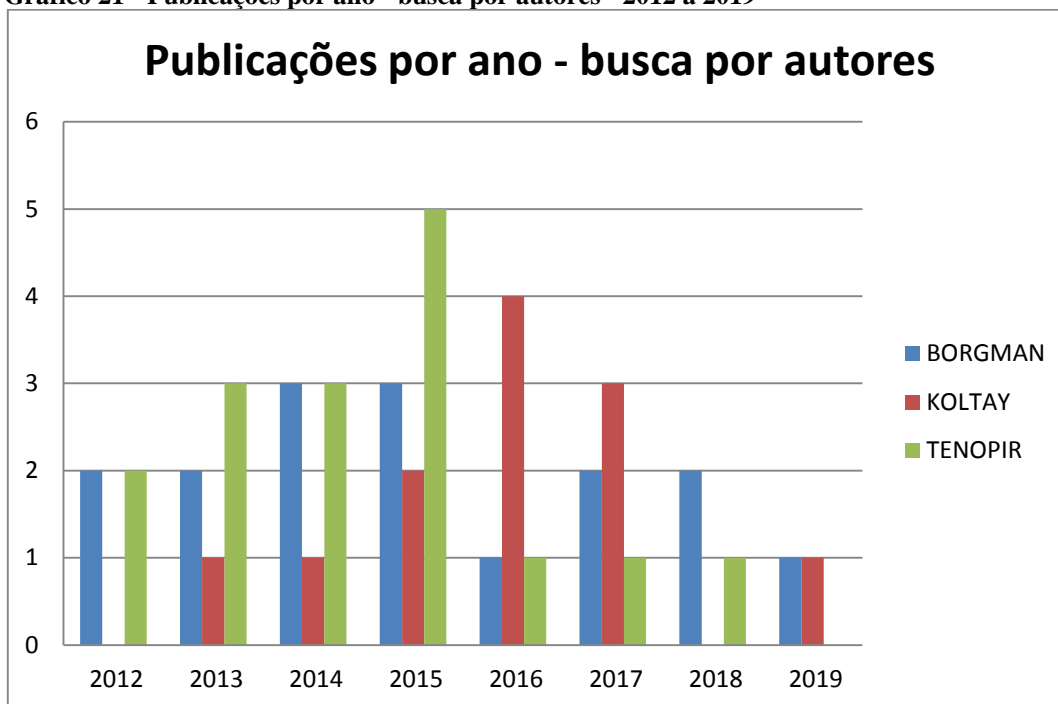
Tabela 18 - Pesquisa por autores – corpus dinâmico de 2012 a 2019

Autores	BORGMAN	KOLTAY	TENOPIR
2012	2	0	2
2013	2	1	3
2014	3	1	3
2015	3	2	5
2016	1	4	1
2017	2	3	1
2018	2	0	1
2019	1	1	0
TOTAL	16	12	16
%	36%	27%	36%
TOTAL Autores	44		

Fonte: Autoria própria (2019).

A evolução de produção dos registros recuperados na estratégia de busca por autores pode ser vista no Gráfico 21.

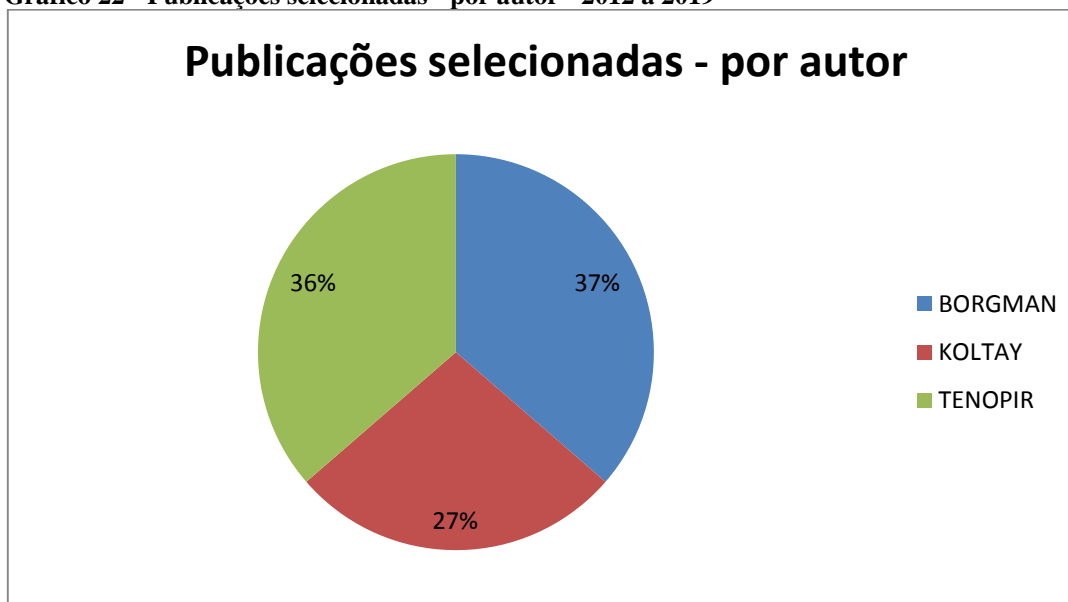
Gráfico 21 - Publicações por ano - busca por autores - 2012 a 2019



Fonte: Autoria própria (2019).

A proporção dos registros selecionados na estratégia de busca por autores pode ser vista no Gráfico 22.

Gráfico 22 - Publicações selecionadas - por autor - 2012 a 2019



Fonte: Autoria própria (2019).

A base de dados WoS oferece uma ferramenta de análise de citações dos artigos. A partir dos registros recuperados nesta fonte, elaborou-se a Tabela 17 para demonstrar a frequência de citação dos autores com maior produção científica. Esta tabela considera apenas as referências listadas na base WoS, sendo excluídas as referências localizadas em outras bases de dados bibliográficas.

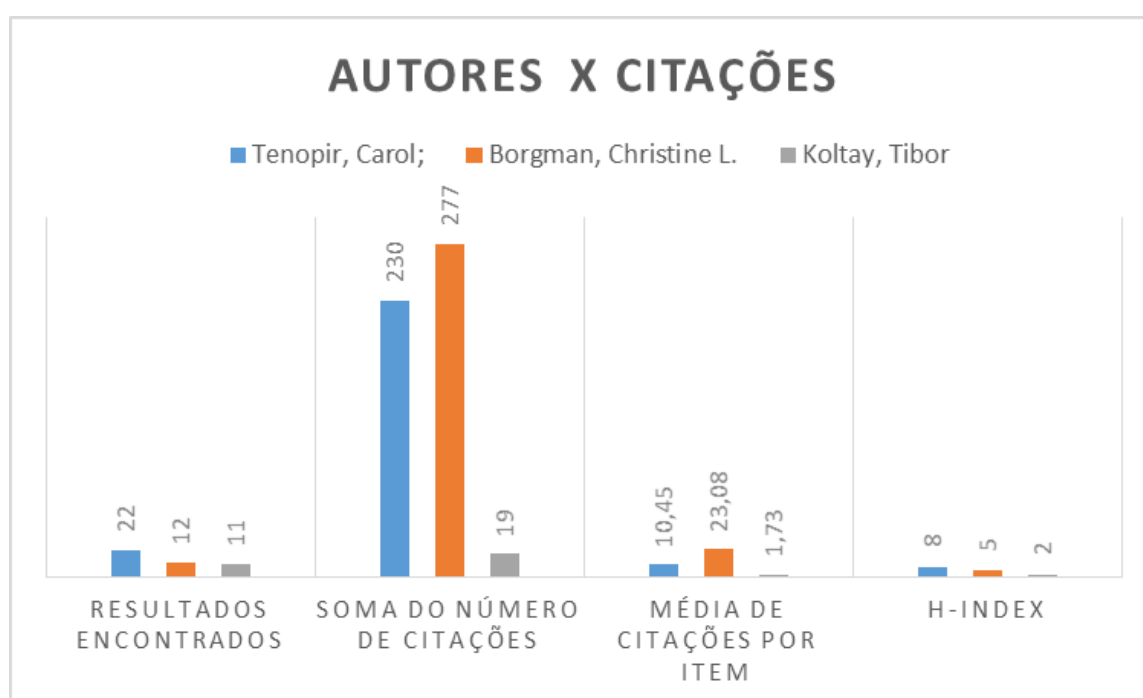
Tabela 19 - Autores e artigos mais citados - resumo

Artigos mais citados	Tenopir, Carol;	Borgman, Christine L.	Koltay, Tibor	TOTAL
Resultados encontrados	22	12	11	45
Soma do número de citações	230	277	19	526
Média de citações por item	10,45	23,08	1,73	
h-index	8	5	2	

Fonte: Adaptado de Web os Science (2018).

A representação gráfica da Tabela 17 pode ser visualizada no Gráfico 23. A autora Borgman, com 12 publicações, possui uma frequência maior de citações (277) que a autora Tenopir (230), cuja produção científica nesta base de dados alcança 22 registros. O autor Koltay possui 11 publicações na base de dados WoS, com 19 citações, porém, seu trabalho está relacionada de forma mais estrita com o letramento em dados, assunto menos abordado na literatura em geral, conforme apresentado no Gráfico 1.

Gráfico 23 - Autores e artigos mais citados - resumo



Fonte: Autoria própria, adaptado de Web os Science (2018).

Quadro 31 - Autores e artigos mais citados - Tenopir

(continua)

Autor-chave	Autores	Título	Título da fonte	Ano da publicação	Título da conferência	Total de citações
Tenopir, Carol;	Michener, William K.; Allard, Suzie; Budden, Amber; Cook, Robert B.; Douglass, Kimberly; Frame, Mike; Kelling, Steve; Koskela, Rebecca; Tenopir, Carol; Vieglais, David A.	Participatory design of DataONE-Enabling cyberinfrastructure for the biological and environmental sciences	ECOLOGICAL INFORMATICS	2012		38
Tenopir, Carol;	Nicholas, David; Watkinson, Anthony; Jamali, Hamid R.; Herman, Eti; Tenopir, Carol; Volentine, Rachel; Allard, Suzie; Levine, Kenneth	Peer review: still king in the digital age	LEARNED PUBLISHING	2015		28
Tenopir, Carol;	Tenopir, Carol; Sandusky, Robert J.; Allard, Suzie; Birch, Ben	Research data management services in academic research libraries and perceptions of librarians	LIBRARY & INFORMATION SCIENCE RESEARCH	2014		25
Tenopir, Carol;	Tenopir, Carol; Dalton, Elizabeth D.; Allard, Suzie; Frame, Mike; Pjesivac, Ivanka; Birch, Ben; Pollock, Danielle; Dorsett, Kristina	Changes in Data Sharing and Data Reuse Practices and Perceptions among Scientists Worldwide	PLOS ONE	2015		25
Tenopir, Carol;	Nicholas, David; Watkinson, Anthony; Volentine, Rachel; Allard, Suzie; Levine, Kenneth; Tenopir, Carol; Herman, Eti	Trust and Authority in Scholarly Communications in the Light of the Digital Transition: setting the scene for a major study	LEARNED PUBLISHING	2014		24
Tenopir, Carol;	Tenopir, Carol; Volentine, Rachel; King, Donald W.	Social media and scholarly reading	ONLINE INFORMATION REVIEW	2013		17
Tenopir, Carol;	Jamali, Hamid R.; Nicholas, David; Watkinson, Anthony; Herman, Eti; Tenopir, Carol; Levine, Kenneth; Allard, Suzie; Christian, Lisa; Volentine, Rachel; Boehm, Reid; Nichols, Frances	How scholars implement trust in their reading, citing and publishing activities: Geographical differences	LIBRARY & INFORMATION SCIENCE RESEARCH	2014		15
Tenopir, Carol;	Watkinson, Anthony; Nicholas, David; Thornley, Clare; Herman, Eti; Jamali, Hamid R.; Volentine, Rachel; Allard, Suzie; Levine, Kenneth; Tenopir, Carol	Changes in the digital scholarly environment and issues of trust: An exploratory, qualitative analysis	INFORMATION PROCESSING & MANAGEMENT	2016		9
Tenopir, Carol;	Volentine, Rachel; Tenopir, Carol	Value of academic reading and value of the library in academics' own words	ASLIB PROCEEDINGS	2013		8

(conclusão)

Autor-chave	Autores	Título	Título da fonte	Ano da publicação	Título da conferência	Total de citações
Tenopir, Carol;	Tenopir, Carol; King, Donald W.; Christian, Lisa; Volentine, Rachel	Scholarly article seeking, reading, and use: a continuing evolution from print to electronic in the sciences and social sciences	LEARNED PUBLISHING	2015		8
Tenopir, Carol;	Tenopir, Carol; Levine, Kenneth; Allard, Suzie; Christian, Lisa; Volentine, Rachel; Boehm, Reid; Nichols, Frances; Nicholas, David; Jamali, Hamid R.; Herman, Eti; Watkinson, Anthony	Trustworthiness and Authority of Scholarly Information in a Digital Age: Results of an International Questionnaire	JOURNAL OF THE ASSOCIATION FOR INFORMATION SCIENCE AND TECHNOLOGY	2016		8
Tenopir, Carol;	Douglass, Kimberly; Allard, Suzie; Tenopir, Carol; Wu, Lei; Frame, Mike	Managing Scientific Data as Public Assets: Data Sharing Practices and Policies Among Full-Time Government Employees	JOURNAL OF THE ASSOCIATION FOR INFORMATION SCIENCE AND TECHNOLOGY	2014		7
Tenopir, Carol;	Tenopir, Carol; Volentine, Rachel; King, Donald W.	Article and book reading patterns of scholars: findings for publishers	LEARNED PUBLISHING	2012		6
Tenopir, Carol;	Thornley, Clare; Watkinson, Anthony; Nicholas, David; Volentine, Rachel; Jamali, Hamid R.; Herman, Eti; Allard, Suzie; Levine, Kenneth J.; Tenopir, Carol	The role of trust and authority in the citation behaviour of researchers	INFORMATION RESEARCH-AN INTERNATIONAL ELECTRONIC JOURNAL	2015		5
Tenopir, Carol;	Tenopir, Carol; Dalton, Elizabeth; Fish, Allison; Christian, Lisa; Jones, Misty; Smith, MacKenzie	What Motivates Authors of Scholarly Articles? The Importance of Journal Attributes and Potential Audience on Publication Choice	PUBLICATIONS	2016		3
Tenopir, Carol;	Tenopir, Carol; Dalton, Elizabeth D.; Christian, Lisa; Jones, Misty K.; McCabe, Mark; Smith, MacKenzie; Fish, Allison	Imagining a Gold Open Access Future: Attitudes, Behaviors, and Funding Scenarios among Authors of Academic Scholarship	COLLEGE & RESEARCH LIBRARIES	2017		2
Tenopir, Carol;	Davis, Miriam L. E. Steiner; Tenopir, Carol; Allard, Suzie; Frame, Michael T.	Facilitating Access to Biodiversity Information: A Survey of Users' Needs and Practices	ENVIRONMENTAL MANAGEMENT	2014		1
Tenopir, Carol;	Forrester, Amy; Bjoerk, Bo-Christer; Tenopir, Carol	New web services that help authors choose journals	LEARNED PUBLISHING	2017		1

Fonte: Autoria própria, adaptado de Web os Science (2018).

Quadro 32 - Autores e artigos mais citados - Borgman

(continua)

Autor-chave	Autores	Título	Título da fonte	Ano da publicação	Título da conferência	Total de citações
Borgman, Christine L.	Borgman, Christine L.	The Conundrum of Sharing Research Data	JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE AND TECHNOLOGY	2012		152
Borgman, Christine L.	Wallis, Jillian C.; Rolando, Elizabeth; Borgman, Christine L.	If We Share Data, Will Anyone Use Them? Data Sharing and Reuse in the Long Tail of Science and Technology	PLOS ONE	2013		61
Borgman, Christine L.	Goodman, Alyssa; Pepe, Alberto; Blocker, Alexander W.; Borgman, Christine L.; Cranmer, Kyle; Crosas, Merce; Di Stefano, Rosanne; Gil, Yolanda; Groth, Paul; Hedstrom, Margaret; Hogg, David W.; Kashyap, Vinay; Mahabal, Ashish; Siemiginowska, Aneta; Slavkovic, Aleksandra	Ten Simple Rules for the Care and Feeding of Scientific Data	PLOS COMPUTATIONAL BIOLOGY	2014		31
Borgman, Christine L.	Borgman, Christine L.; Wallis, Jillian C.; Mayernik, Matthew S.	Who's Got the Data? Interdependencies in Science and Technology Collaborations	COMPUTER SUPPORTED COOPERATIVE WORK-THE JOURNAL OF COLLABORATIVE COMPUTING	2012		17
Borgman, Christine L.	Mayernik, Matthew S.; Wallis, Jillian C.; Borgman, Christine L.	Unearthing the Infrastructure: Humans and Sensors in Field-Based Scientific Research	COMPUTER SUPPORTED COOPERATIVE WORK-THE JOURNAL OF COLLABORATIVE COMPUTING	2013		9
Borgman, Christine L.	Darch, Peter T.; Borgman, Christine L.; Traweek, Sharon; Cummings, Rebekah L.; Wallis, Jillian C.; Sands, Ashley E.	What lies beneath?: Knowledge infrastructures in the subseafloor biosphere and beyond	INTERNATIONAL JOURNAL ON DIGITAL LIBRARIES	2015		3
Borgman, Christine L.	Borgman, Christine L.; Darch, Peter T.; Sands, Ashley E.; Pasquetto, Irene V.; Golshan, Milena S.; Wallis, Jillian C.; Traweek, Sharon	Knowledge infrastructures in science: data, diversity, and digital libraries	INTERNATIONAL JOURNAL ON DIGITAL LIBRARIES	2015		2

(conclusão)

Autor-chave	Autores	Título	Título da fonte	Ano da publicação	Título da conferência	Total de citações
Borgman, Christine L.	Pasquetto, Irene V.; Sands, Ashley E.; Darch, Peter T.; Borgman, Christine L.	Open Data in Scientific Settings: From Policy to Practice	34TH ANNUAL CHI CONFERENCE ON HUMAN FACTORS IN COMPUTING SYSTEMS, CHI 2016	2016	34th Annual CHI Conference on Human Factors in Computing Systems (CHI4GOOD)	1
Borgman, Christine L.	Borgman, Christine L.; Darch, Peter T.; Sands, Ashley E.; Wallis, Jillian C.; Traweek, Sharon	The Ups and Downs of Knowledge Infrastructures in Science: Implications for Data Management	2014 IEEE/ACM JOINT CONFERENCE ON DIGITAL LIBRARIES (JCDL)	2014	14th IEEE/ACM Joint Conference on Digital Libraries (JCDL) / 18th International Conference on Theory and Practice of Digital Libraries (TPDL)	1

Fonte: Autoria própria, adaptado de Web os Science (2018).

Quadro 33 - Autores e artigos mais citados - Koltay

Autor-chave	Autores	Título	Título da fonte	Ano da publicação	Título da conferência	Total de citações
Koltay, Tibor	Koltay, Tibor	Data literacy: in search of a name and identity	JOURNAL OF DOCUMENTATION	2015		8
Koltay, Tibor	Koltay, Tibor; Spiranec, Sonja; Karvalics, Laszlo Z.	The Shift of Information Literacy Towards Research 2.0	JOURNAL OF ACADEMIC LIBRARIANSHIP	2015		5
Koltay, Tibor	Koltay, Tibor	Data literacy for researchers and data librarians	JOURNAL OF LIBRARIANSHIP AND INFORMATION SCIENCE	2017		2
Koltay, Tibor	Todorova, Tania Yordanova; Kurbanoglu, Serap; Boustany, Joumana; Dogan, Guleda; Saunders, Laura; Horvat, Aleksandra; Terra, Ana Lucia; Landoy, Ane; Repanovici, Angela; Morrison, Chris; Vanderkast, Egbert J. Sanchez; Secker, Jane; Rudzioniene, Jurgita; Kortelainen, Terttu; Koltay, Tibor	Information professionals and copyright literacy: a multinational study	LIBRARY MANAGEMENT	2017		1
Koltay, Tibor	Koltay, Tibor	Are you ready? Tasks and roles for academic libraries in supporting Research 2.0	NEW LIBRARY WORLD	2016		1
Koltay, Tibor	Koltay, Tibor	Library and information science and the digital humanities Perceived and real strengths and weaknesses	JOURNAL OF DOCUMENTATION	2016		1
Koltay, Tibor	Koltay, Tibor; Varga, Katalin	Conceptions, Ideas, What Else? Information Literacy in Hungary	WORLDWIDE COMMONALITIES AND CHALLENGES IN INFORMATION LITERACY RESEARCH AND PRACTICE	2013	1st European Conference on Information Literacy (ECIL)	1

Fonte: Autoria própria, adaptado de Web os Science (2018)

APÊNDICE C – Seleção de artigos para revisão sistemática – 2018

As etapas da revisão sistemática que estão detalhadas na Tabela 20 são: Etapa 01 – Exportados da recuperação na busca; Etapa 02 – Eliminação dos duplicados; Etapa 03 – Selecionados trabalhos potenciais após leitura do título e resumo; Etapa 04 - Selecionados para revisão - Aderentes à pesquisa; Etapa 05 - Texto completo disponíveis selecionados para revisão.

Tabela 20 - Etapas de seleção de artigos para revisão sistemática

Estratégia de busca	Etapas de Pesquisa	Portal da Capes		Ebsco Host		Web of Science		Total
		Número	%	Número	%	Número	%	
Combinação C	Etapa 01	20	12%	119	72%	26	16%	165
	Etapa 02	18	12%	108	73%	21	14%	147
	Etapa 03	13	10%	103	79%	14	11%	130
	Etapa 04	9	14%	45	70%	10	16%	64
	Etapa 05	6	18%	23	70%	4	12%	33
Combinação E	Etapa 01	5	9%	22	42%	26	49%	53
	Etapa 02	4	8%	22	42%	26	50%	52
	Etapa 03	4	10%	20	50%	16	40%	40
	Etapa 04	2	10%	11	52%	8	38%	21
	Etapa 05	2	17%	6	50%	4	33%	12
BORGMAN	Etapa 01	15	35%	16	37%	12	28%	43
	Etapa 02	14	33%	16	38%	12	29%	42
	Etapa 03	12	44%	11	41%	4	15%	27
	Etapa 04	9	56%	6	38%	1	6%	16
	Etapa 05	7	50%	6	43%	1	7%	14
KOLTAY	Etapa 01	20	30%	35	53%	11	17%	66
	Etapa 02	20	34%	28	48%	10	17%	58
	Etapa 03	20	59%	10	29%	4	12%	34
	Etapa 04	11	92%	1	8%	0	0%	12
	Etapa 05	9	90%	1	10%	0	0%	10
TENOPIR	Etapa 01	32	23%	83	61%	22	16%	137
	Etapa 02	32	30%	53	50%	22	21%	107
	Etapa 03	12	26%	34	74%	0	0%	46
	Etapa 04	6	38%	10	63%	0	0%	16
	Etapa 05	5	42%	7	58%	0	0%	12

Fonte: Autoria própria (2017). Link para planilha completa: <https://goo.gl/dGBpgH>
Listada na Tabela 10.

APÊNDICE D – Plano de gestão de dados (*Data Management Plan –DMP*)

LETRAMENTO INFORMACIONAL PARA REÚSO DE DADOS NAS CIÊNCIAS SOCIAIS EM AMBIENTES VIRTUAIS DE PESQUISA: proposta de requisitos e competências

Nome do pesquisador principal	Janete Saldanha Bach Estevão
E-mail:	janeteest@gmail.com
ORCID	https://orcid.org/0000-0001-9011-5837
Curriculum Lattes:	http://lattes.cnpq.br/2190313867279271
Grupo de Pesquisa	Território - Redes, Políticas, Tecnologia e Desenvolvimento - UTFPR. http://dgp.cnpq.br/dgp/espelhogrupo/7433465531776870
Agência de fomento	-
Instituição	UTFPR – Universidade Tecnológica Federal do Paraná
Supervisão / Orientação	Profa. Dra. Faimara do Rocio Strauss
Duração	Início: 01-03-2017 Fim: 28-02-2020
Data de elaboração / Versão	02-01-2018 / Versão 01
Repositório de armazenamento de longo prazo	Base de Dados Científicos da Universidade Federal do Paraná (BDC/UFPR): https://bdc.c3sl.ufpr.br
DOI para os dados deste PGD	Aguardando

Sumário	
1. Introdução	352
2. Identificação da pesquisa:	352
2.1 Título da pesquisa:	352
2.2 Objetivos da pesquisa:	352
3. Descrição dos dados:	352
3.1 Tipos de dados	352
3.2 Organização e Documentação	353
3.3 Metadados	353
4. Proteção / Privacidade	355
5. Armazenamento, Backup, Segurança	355
5.1 Armazenamento e Backup	355
5.2 Capacidade de dados e volume	356
6 Acesso e Preservação a Longo Prazo	356
7 Compartilhamento de Dados e Disseminação	356
8 Políticas para reutilização	357
8.1 Propriedade intelectual	357
9 Citação e Licença	357
Referências	357

1. Introdução

Esse plano de gestão de dados (DMP) faz parte de um projeto de pesquisa de doutorado que foi executado em 33 meses (março/2017 a dezembro/2019). Esse projeto de pesquisa investiga o acesso aberto e o reúso dos dados de pesquisa já disponibilizados, prática que tem como base as concepções da Ciência Aberta.

1.1 Descreva as funções e responsabilidades

Janete Saldanha Bach Estevão será responsável pela implementação, pelo desenvolvimento e pela manutenção deste conjunto de dados.

1.2 Quem terá a responsabilidade pelos dados ao longo do tempo?

Janete Saldanha Bach Estevão será a responsável pela gestão de dados e pelo monitoramento deste Plano de Gestão de Dados.

Identificação da pesquisa:

2. Identificação da pesquisa:

2.1 Título da pesquisa:

Letramento informacional para reúso de dados nas Ciências Sociais em ambientes virtuais de pesquisa: proposta de requisitos e competências.

2.2 Objetivos da pesquisa:

Propor, a partir de elementos da Teoria Ator-Rede, requisitos para um ambiente virtual em letramento informacional de dados, com as competências mínimas para o reúso de dados em Ciências Sociais. Para isso, utilizará os quadros conceituais (*frameworks*) de letramento informacional em dados, as dificuldades no reúso dados de pesquisa, por meio de dados secundários (*datasets*) e os recursos de aprendizado *online*, disponíveis nos ambientes virtuais de aprendizagem de letramento informacional em dados.

3. Descrição dos dados:

Tipos de dados

- a) **Dados secundários / *frameworks*:** Os dados a serem utilizados serão os conteúdos da literatura em formato de *frameworks* em letramento informacional de dados, principalmente textos em formato pdf. Esses textos serão processados em documentos de editor de texto ou planilha eletrônica para serem processados no *software* de análise de conteúdo NVivo. Os resultados serão dados qualitativos exportados para arquivos eletrônicos em formatos não proprietários para posterior disponibilização.
- b) **Dados secundários / recursos *online*:** Outro conjunto de dados qualitativos proveniente dos recursos *online* disponíveis para o letramento em dados são arquivos eletrônicos em formato pdf, editor de texto, planilha eletrônica, formato texto, vídeos e conteúdo html. Esses arquivos serão processados igualmente no *software* NVivo e

resultados dessas análises serão dados qualitativos exportados para arquivos eletrônicos em formatos não proprietários para posterior disponibilização.

- c) **Dados secundários/surveys:** Os dados quantitativos são provenientes de levantamentos (*surveys*) com pesquisadores sobre a prática de compartilhamento e reúso de dados de pesquisa. Foram entrevistados nos 3 conjuntos de dados reutilizados um total de 5.198 pesquisadores. Esses dados estão disponíveis em repositórios de dados em acesso aberto. Serão processados pelo *software* NVivo, que é usado para a maioria das análises secundárias. Não há necessidade de conversão pois o NVivo suporta diversos tipos de formatos, como por exemplo: texto, planilha, Pdfs, vídeos, *websites*, entre outros. Procuraremos desenvolver funções de análise documentadas que executem as análises nos dados primários e produzam os arquivos de dados secundários junto com as figuras, quando necessário. Arquivos de dados brutos e PDF conterão tabelas detalhando colunas e linhas, identificadores e conteúdo, descrição de formatos e outros elementos para a interpretação dos dados.

3.1 Organização e Documentação

Os arquivos fonte dos *frameworks* de letramento informacional em dados serão armazenados em PDF, com os respectivos metadados em uma pasta pública do gerenciador de referências Zotero. Também serão armazenados em pasta virtual em nuvem, nomeados pelo último sobrenome(s) do(s) autor(res) e o ano de publicação, da mesma forma que se faz a citação no texto. Os arquivos processados, ou seja, os *frameworks* transferidos para os arquivos em formato texto ou planilha eletrônica serão armazenados em pasta virtual em nuvem e posteriormente disponíveis em acesso aberto, desde que a fonte em que este *framework* foi publicado esteja em acesso aberto também.

Os arquivos dos recursos *online* disponíveis para o letramento em dados serão igualmente organizados no gerenciador de referências Zotero, com os metadados e os anexos de cada conteúdo, seguindo a mesma atribuição de licença de acesso e uso Creative Commons. Uma cópia do que for permitido para *download* será armazenada em pasta virtual na nuvem. Esses arquivos serão mantidos apenas durante a análise de conteúdo e processamento dos dados. Posteriormente, links indicarão o acesso ao conteúdo. Armazenar os mesmos em repositório institucional seria duplicação de informação e uso desnecessário de espaço em disco.

Os arquivos eletrônicos dos levantamentos (*surveys*) serão analisados e os resultados desta análise secundária será disponibilizada em formato bruto com uso de planilhas eletrônicas não de formato proprietário. Esses arquivos processados serão posteriormente disponibilizados no repositório de dados. Os arquivos originais de dados secundários (*datasets*) não serão submetidos novamente em repositórios pois já se encontram em acesso aberto.

3.2 Metadados

Metadados serão fornecidos de acordo com o padrão Dublin Core para cada arquivo processado. Os metadados incluem o nome do arquivo para os resultados, localização, proprietário, tempo de criação, versão do modelo, entradas, informações de *hardware* e *software*, etc. O nome do arquivo de saída aparecerá como Nome_Versão_Data (AAAAMMDD). Os metadados dos arquivos primários serão obtidos automaticamente pelo *software* Zotero. Aqueles que não forem possíveis ser extraídos automaticamente, serão completados pela pesquisadora no formato de referências Bibliográfica da ABNT. Contudo, serão documentadas informações sobre o contexto, conteúdo, qualidade, proveniência e / ou acessibilidade dos dados utilizados. Também serão documentadas informações sobre autores, datas e breves descrições de PDFs primários, conjuntos de dados secundários e recursos de

aprendizagem. Esses dados e informações ficam disponíveis na documentação que acompanha os conjuntos de dados.

a) Os dados utilizados como *frameworks* serão:

- **Framework 01** Baseado no *framework* de letramento informacional da ACRL (2000) Carlson *et al.* (2011);
- **Framework 02** Baseado no *framework* de letramento informacional da ACRL (2000) Calzada Prado; Marzal (2013);
- **Framework 03** Facilitate Open Science Training For European Research. Foster (2015);
- **Framework 04** Vitae Researcher Development Framework - Vitae (2011);
- **Framework 05** Data Information Literacy – DIL Carlson; Johnston (2015).
- **Framework 06** University of Minnesota *Data management course*; Jeffryes; Johnston (2013).

b) Os dados utilizados como recursos *online* serão (ver Quadro 4):

- Bodleian Data Library – University of Oxford (UNIVERSITY OF OXFORD, 2018-2019).
 - <https://libguides.bodleian.ox.ac.uk/workshops/rdm>
- Consortium of European Social Science Data Archives (CESSDA) e European Research Infrastructure Consortium (ERIC). (CONSORTIUM OF EUROPEAN SOCIAL SCIENCE DATA ARCHIVES, 2018).
 - <https://www.cessda.eu/Research-Infrastructure/Training/Expert-Tour-Guide-on-Data-Management>
- Data Management Course (JOHNSTON; JEFFRYES, 2012).
 - <https://libguides.bodleian.ox.ac.uk/workshops/rdm>
- Data Observation Network for Earth – (DataONE) (DATAONE, 2012).
 - www.dataone.org/education-modules.
- Digital Curation Center - DC 101 materials: Access and Reuse (DIGITAL CURATION CENTER, 2014-2019).
 - <http://www.dcc.ac.uk/training/train-the-trainer/dc-101-training-materials>
- Essentials 4 Data Support (RESEARCH DATA NETHERLANDS, *online*).
 - <http://datasupport.researchdata.nl/en/>
- Facilitate Open Science Training for European Research (FOSTER, 2015).
 - www.fosteropenscience.eu;
- Managing and Sharing Research Data handbook (CORTI *et al.*, 2014).
 - www.ukdataservice.ac.uk/manage-data/handbook

- Research Data and Management Training - MANTRA (UNIVERSITY OF EDINBURGH; EDINA Data Library, 2017).
 - <https://mantra.edina.ac.uk/>
 - New England Collaborative Data Management Curriculum (NECDMC) (NECDMC, 2012).
 - <https://library.umassmed.edu/resources/necdmc/index>
 - Research Data Management and Sharing - Coursera (CURATING RESEARCH ASSETS AND DATA USING LIFECYCLE EDUCATION (CRADLE); UNIVERSITY OF NORTH CAROLINA; UNIVERSITY OF EDINBURGH, 2017).
 - www.coursera.org/learn/research-data-management-and-sharing
- c) Os dados utilizados levantamentos (*surveys*) de dados secundários sobre pesquisadores serão:
- BERGHMANS, Stephane; COUSIJN, Helena; DEAKIN, Gemma, DEAKIN, Gemma; MEIJER, Ingeborg; MULLIGAN, Adrian; PLUME, Andrew; RIJCKE, Sarah de; RUSHFORTH, Alex; TATUM, Clifford; LEEUWEN, Thed van; WALTMAN, Ludo. **Open Data: the researcher perspective - survey and case studies.** [Dataset]. 2017. Mendeley Data, versão 1, 4 Abr. 2017. Disponível em: <DOI: 10.17632/bwrnfb4bvh.> Acesso em: 25 abr. 2017.
 - KIM, Youngseek. **Scientific Data Reuse Survey** [Dataset]. Inter-university Consortium for Political and Social Research [distribuidor]. 2017. Disponível em: <<http://doi.org/10.3886/E100404V1>>. Acesso em: 25 abr. 2017.
 - KIM, Youngseek. **Scientists' Data Sharing Behaviors** [Dataset]. Inter-university Consortium for Political and Social Research [distribuidor]. 2016. Disponível em: <<http://doi.org/10.3886/E100087V7>>. Acesso em: 25 abr. 2017.

4. Proteção / Privacidade

Os dados a serem adquiridos a partir da pesquisa proposta não envolvem sujeitos humanos ou animais, mas pesquisa de bibliografia e de dados secundários (*surveys*). Para os levantamentos (*surveys*), em caso dos dados revelarem alguma informação sensível, como nomes ou conteúdo que revele identidades e viole a privacidade de público entrevistado ou pesquisado, esse conteúdo será suprimido e, caso não seja possível suprimir, será extraído da análise. Porém, como são dados secundários, já passaram por um crivo de avaliação e uma autorização do comitê de ética na instituição em que a pesquisa foi originalmente realizada, o que confere mais segurança de proteção aos dados sensíveis.

5. Armazenamento, Backup, Segurança

5.1 Armazenamento e Backup

Dados secundários e processados serão armazenados em pasta virtual na nuvem, cujo backup será em disco rígido, apenas para uma eventual perda de acesso. A senha para acesso a esta pasta virtual é elaborada a partir de critérios de consistência e um segundo fator de autenticação é requerido, para evitar ação de *hackers*.

5.2 Capacidade de dados e volume

O repositório (BDC/UFPR) tem capacidade para armazenar até 2Gb para cada conjunto de dados. Até o momento os dados coletados são distribuídos da seguinte forma, totalizam 834 MB:

- *Frameworks*: 38,5 MB
- *Surveys*: 38,8 MB
- Recursos *online* 757 MB

Espera-se no final gerar cerca de 1GB considerando os dados processados.

Posteriormente os dados submetidos ao armazenamento de longo prazo na Base de Dados Científicos da Universidade Federal do Paraná (BDC/UFPR), disponível no *link* <https://bdc.c3sl.ufpr.br>, em um período não inferior a 2 anos após completar o projeto. A BDC/UFPR fornece armazenamento centralizado, com backup e acesso à rede para projetos de pesquisa sem custo. Mantém um ambiente de tecnologia de informação distribuída, com aspectos centrais e locais de planejamento e controle gerais.

6 Acesso e Preservação a Longo Prazo

Dois anos após a publicação, ou após a conclusão da pesquisa, os dados processados serão disponibilizados para a comunidade. O acesso se dará via sítio da Base de Dados Científicos da Universidade Federal do Paraná (BDC/UFPR) (<https://bdc.c3sl.ufpr.br>), que não impõe qualquer barreira a pesquisa e descoberta de dados, como cadastros, vínculos institucionais, entre outros. Os dados de pesquisa para este estudo serão armazenados na base de dados científicos que é um dos serviços do repositório institucional da UFPR. Esse repositório é muito bem posicionado no *ranking webometrics*²⁸⁰, sendo o 9. No âmbito nacional dentre 1400 registrados. O serviço de armazenamento de dados científicos está registrado no indexador re3data.org²⁸¹, o que favorece a descoberta dos conjuntos de dados ali depositados.

O acesso exige uso da rede de internet e pode ser feito em qualquer navegador. Os arquivos serão acompanhados de documentação que atribui significado aos dados, em formato de arquivo eletrônico não proprietário. *Softwares* a serem utilizados para abrir os arquivos serão indicados.

7 Compartilhamento de Dados e Disseminação

Os dados serão divulgados por publicação em revistas acadêmicas de artigos e/ou em revistas de publicação de dados (*Data papers*). Os dados em artigos de periódicos e informações de apoio serão apresentados em formato padrão. Essas informações incluirão descrições dos procedimentos de pesquisa, descrições numéricas, textuais e visuais dos resultados após o processamento do dados.

Os dados processados serão submetidos em estado bruto, sob uma licença Creative Commons CC-BY, e estarão disponíveis para *download* em até dois anos após o término do projeto no repositório (BDC/UFPR). Termos e condições (por exemplo, autoria,

²⁸⁰ <http://www.webometrics.info/en>

²⁸¹ <http://re3data.org/>

reconhecimento, outra compensação) estarão expressas nessa licença. Dada a natureza dos dados, não há questões de privacidade, legais, regulatórias ou éticas associadas à disseminação desses dados. Tanto quanto possível, os resultados serão publicados em periódicos de acesso aberto, caso seja necessário, será preciso apoio financeiro para publicação de acesso aberto da Instituição de Pesquisa.

8 Políticas para reutilização

Não haverá quaisquer restrições para reuso de dados, pois estarão sob a Creative Commons CC-BY, apenas a citação se fará necessária, com indicação de DOI.

8.1 Propriedade intelectual

A pesquisadora principal reterá os principais direitos legais sobre a propriedade intelectual desenvolvidos sob esta concessão, quesito que está em conformidade com a “Diretrizes da Base de Dados Científicos da Universidade Federal do Paraná” (http://portal.ufpr.br/documentos/BDC/diretrizes_BDC.pdf).

9 Citação e Licença

Como citar:

ESTEVÃO, Janete Saldanha Bach. **Plano de Gestão de Dados para Letramento informacional para reuso de dados nas Ciências Sociais em Ambientes virtuais de pesquisa: requisitos e competências essenciais**. Curitiba, 2019. 9 p.

Licença de uso

Plano de Gestão de Dados para Letramento informacional para reuso de dados nas Ciências Sociais em Ambientes virtuais de pesquisa: requisitos e competências essenciais, by Janete Saldanha Bach Estevão is licensed under a Creative Commons Attribution 3.0 International Licence BY-NC-SA.

Referências

BERGHMANS, Stephane; COUSIJN, Helena; DEAKIN, Gemma, DEAKIN, Gemma; MEIJER, Ingeborg; MULLIGAN, Adrian; PLUME, Andrew; RIJCKE, Sarah de; RUSHFORTH, Alex; TATUM, Clifford; LEEUWEN, Thed van; WALTMAN, Ludo. **Open Data: the researcher perspective - survey and case studies**. [Dataset]. 2017. Mendeley Data, versão 1, 4 Abr. 2017. Disponível em: DOI: 10.17632/bwrnfb4bvh. Acesso em: 25 abr. 2017.

CALZADA PRADO, Javier; MARZAL, Miguel Ángel. Incorporating Data Literacy into Information Literacy Programs: Core Competencies and Contents. **Libri: International Journal of Libraries & Information Services**. v. 63, n.2, p. 123–34. 2013. Disponível em: <https://doi.org/10.1515/libri-2013-0010>. Acesso em: 02 jul. 2017.

CARLSON, Jacob; FOSMIRE, Michael; MILLER, C.C.; NELSON, Megan Sapp. Determining Data Information Literacy Needs: A Study of Students and Research Faculty. **Portal-Libraries and the Academy**, v. 11, n. 2, p. 629-657, abr. 2011. Disponível em: <https://goo.gl/yh4TBD>. Acesso em: 12 dez. 2017.

CARLSON, Jake; JOHNSTON, Lisa R., orgs. **Data Information Literacy: Librarians, Data, and the Education of a New Generation of Researchers**. Purdue Information Literacy Handbooks. West Lafayette, Indiana: Purdue University Press, 2015. Disponível em: <https://goo.gl/riC4Es>. Acesso em: 12 dez. 2017.

FACILITATE OPEN SCIENCE TRAINING FOR EUROPEAN RESEARCH. Open Science Learning Objectives, 2015. FOSTER Consortium Members, DOI: 10.5281/zenodo.15603. Disponível em: <https://goo.gl/Lkm1XT>. Acesso em: 20 jul.2016.

KIM, Youngseek. **Scientific Data Reuse Survey** [Dataset]. Inter-university Consortium for Political and Social Research [distribuidor]. 2017. Disponível em: <http://doi.org/10.3886/E100404V1>. Acesso em: 25 abr. 2017.

KIM, Youngseek. **Scientists' Data Sharing Behaviors** [Dataset]. Inter-university Consortium for Political and Social Research [distribuidor]. 2016. <http://doi.org/10.3886/E100087V7>.

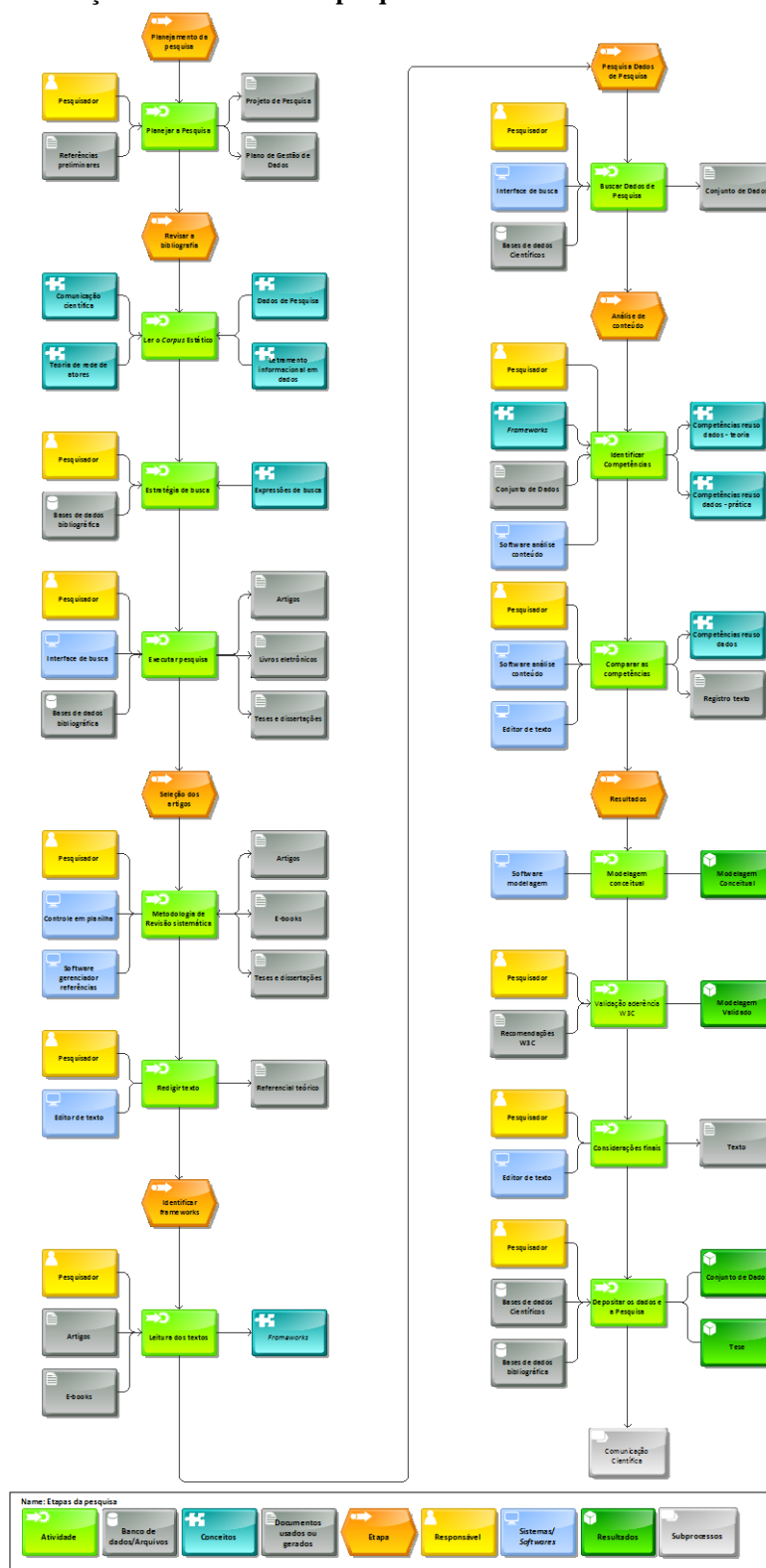
LAMAR SOUTTER LIBRARY, University of Massachusetts Medical School. **Simplified Data Management Plan: New England Collaborative Data Management Curriculum**. Disponível em: <http://library.umassmed.edu/necdmc>. Acesso em: 20 jul. 2018.

UNIVERSIDADE FEDERAL DO PARANÁ. **Plano de Gestão de Dados (PGD)**. Disponível em: <https://goo.gl/aWkoMT>. Acesso em: 20 jul. 2018.

VITAE. **The Vitae Researcher Development Framework**. 2018, *online*. Disponível em: <https://goo.gl/XiXfVb>. Acesso em: 05 jan. 2018.

APÊNDICE E – Macro etapas de pesquisa - fluxo da execução das atividades

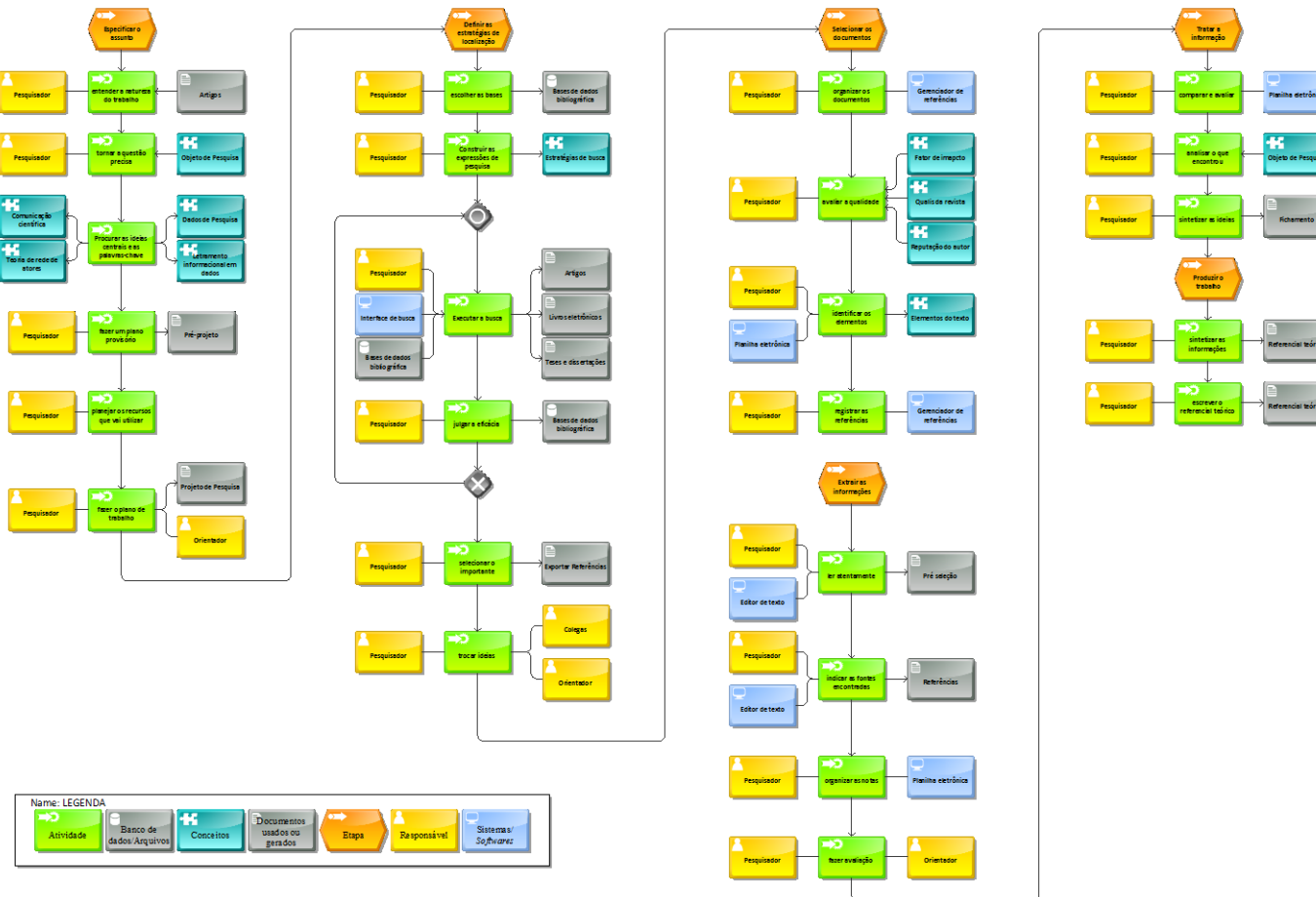
Figura 41 - Fluxo da execução das atividades da pesquisa - 2017-2018



Fonte: Autoria própria (2018).

APÊNDICE F – Revisão sistemática - fluxo da execução das atividades

Figura 42 - Fluxo da execução das atividades de revisão sistemática da literatura



Fonte: Autoria própria (2018).

APÊNDICE G – *Frameworks* das competências do letramento em dados

Quadro 34 - *Framework 01: competências do letramento em dados Carlson et al.(2011).*

(continua)

<i>Framework 01</i>	Competência	Conteúdo
1. Introdução aos bancos de dados e formatos de dados	1.1 Compreender Conceito de bancos de dados relacionais	Entende o conceito de bancos de dados relacionais, como consultar esses bancos de dados
	1.2 Compreender Formatos e tipos de dados padrão por disciplina	Se torna familiarizado com formatos e tipos de dados padrão para sua disciplina
	1.3 Compreender Formatos e tipos de dados para pesquisa	Compreende quais formatos e tipos de dados são apropriados para diferentes questões de pesquisa
2. Descoberta e Aquisição de Dados	2.1 Utilizar repositórios de dados	Localiza e utiliza repositórios de dados disciplinares
	2.2 Importar e converter dados	Identifica não apenas fontes de dados apropriadas, mas também importa dados e os converte quando necessário, para que possam ser usados por ferramentas de processamento downstream
3. Gestão e Organização de Dados	3.1 Compreender o Ciclo de vida dos dados	Compreende o ciclo de vida dos dados
	3.2 Desenvolver Planos de gestão de dados	desenvolve planos de gestão de dados
	3.3 Relacionar os dados	acompanha a relação de subconjuntos ou dados processados para os conjuntos de dados originais
	3.4 Documentar os dados	Cria procedimentos operacionais padrão para gestão de dados e documentação
4. Conversão de Dados e Interoperabilidade	4.1 Migrar de dados	proficiente na migração de dados de um formato para outro
	4.2 Riscos em migração de dados	Compreende os riscos e perda potencial ou corrupção de informações causadas pela mudança de formatos de dados
	4.3 Garantir a qualidade	Entende o Garantia da Qualidade
	4.4 Solucionar corrupção de dados	Reconhece e resolve qualquer aparente artefato, incompletude ou corrupção de conjuntos de dados.
	4.5 Metadados de conjuntos de dados.	Utiliza metadados para facilitar a compreensão de possíveis problemas com conjuntos de dados.

(continua)

<i>Framework 01</i>	Competência	Conteúdo
5. Metadados	5.1 Descrever dos metadados	Compreende a lógica dos metadados e anota proficientemente e descreve os dados para que possam ser entendido e usado por eles mesmos, outros em seu grupo de trabalho e usuários externos.
	5.2 Compreender os metadados	Desenvolve a capacidade de ler e interpretar metadados de fontes disciplinares externas
	5.3 Compreender as ontologias	Compreende a estrutura e propósito das ontologias em facilitar o melhor compartilhamento de dados
6. Curadoria de dados e reutilização	6.1 Valorar os dados	Reconhece que os dados podem ter valor além do seu propósito original, validar pesquisas ou usar por outros
	6.2 Valorar a curadoria de dados	Entende que a curadoria de dados é um empreendimento complexo, muitas vezes dispendioso, vital para a pesquisa eletrônica conduzida pela comunidade
	6.3 Preparar dos dados	Reconhece que os dados devem ser preparados para sua eventual curadoria na sua criação e durante todo o seu ciclo de vida
	6.4 Habilitar a curadoria de dados	Articula o planejamento e as ações necessárias para habilitar a curadoria de dados
7. Culturas da Prática	7.1 Alinhar práticas	Reconhece as práticas, valores e normas de seu campo, disciplina ou sub-disciplina escolhidos
	7.2 Relacionar práticas	Relacionam-se a gestão, compartilhamento, curadoria e preservação de dados
	7.3 Reconhecer padrões de dados	Reconhece padrões de dados relevantes do seu campo (metadados, qualidade, formatação, etc.) e compreende como esses padrões são aplicados
8. Preservação de Dados	8.1 Valorar a preservação de dados	Reconhece os benefícios e custos da preservação de dados
	8.2 Compreende a preservação de dados	Compreende a tecnologia, recursos e componentes organizacionais de preservação de dados
	8.3 Utilizar as melhores práticas de preservação	Utiliza as melhores práticas de preservação apropriadas para o valor e a reprodutibilidade dos dados

(conclusão)

<i>Framework 01</i>	Competência	Conteúdo
9. Análise de dados	9.1 Familiarizar-se com as ferramentas análise de dados	Familiariza-se com as ferramentas básicas de análise de sua disciplina
	9.2 Utilizar o fluxo de trabalho apropriado	Utiliza fluxo de trabalho apropriado ferramentas de gestão para automatizar a análise repetitiva de dados
10. Visualização de dados	10.1 Usar ferramentas de visualização de dados	Proficientemente usa ferramentas básicas de visualização de disciplina
	10.2 Evitar representações enganosas ou ambíguas	Evitar representações enganosas ou ambíguas ao apresentar dados
	10.3 Compreende as vantagens de diferentes tipos de visualização	Compreende as vantagens de diferentes tipos de visualização, por exemplo, mapas, gráficos, animações ou vídeos, ao exibir dados
11. Ética, incluindo citação de dados	11.1 Compreende a propriedade intelectual	Desenvolve uma compreensão da propriedade intelectual, privacidade e questões de confidencialidade, e o ethos da sua disciplina quando se trata de compartilhar dados
	11.2 Reconhece dados de fontes externas	Reconhece apropriadamente dados de fontes externas
	11.3 Disponibilizar dados em formatos padrão	Efeito de disponibilizar dados em formatos padrão para facilitar o uso posterior

**Fonte: Adaptado de Carlson *et al.*(2011).
Listado no Quadro 3.**

Quadro 35 - Framework 02: competências do letramento em dados Calzada Prado e Marzal (2013)

(continua)

<i>Framework 02</i>		Competência	Conteúdos
1 Entendendo dados.	1.1. O que são dados?	saber o que é significava por dados e estar ciente dos vários possíveis tipos de dados	definição de dados; Tipos de dados (dependendo em origem, formato, licença de uso e assim por diante).
	1.2. Dados na sociedade	estar cientes do papel de dados na sociedade, como eles são gerados e por quem, e suas possíveis aplicações, bem como as implicações de seu uso.	produtores e consumidores de dados; Dados ciclo da vida; Aplicações de dados: seu impacto na ciência e sociedade; Direitos autorais e licenças influenciando reutilização de dados
2. Encontrar e / ou obter dados.	2.1. Fontes de dados.	estar cientes do possível fontes de dados, ser capaz de avaliá-las e selecione os mais relevantes para um informativo necessidade ou um determinado problema.	fontes de dados; Critérios para avaliar dados fontes.
	2.2. Obtendo dados.	ser capazes de detectar quando um determinado problema ou necessidade não pode ser (totalmente ou parcialmente) resolvido com os dados existentes e, como apropriado, realizar pesquisas para obter novas dados.	Principais métodos de pesquisa para obter dados originais.
3. Leitura, interpretação e avaliação de dados.	3.1. Ler e interpretar dados.	estar cientes dos vários formulários em que os dados podem ser apresentados (escritos, numérica ou gráfica), e seus respectivos convenções, e ser capaz de interpretá-las.	maneiras de apresentar e representar dados.
	3.2. Avaliando dados.	ser capazes de avaliar dados criticamente.	Critérios de avaliação de dados (incluindo autoria, método de obtenção e análise dados, comparabilidade, inferência e resumos de dados).
4. Gerenciando dados.	4.1. Coleta e gestão de dados e metadados	estar cientes do precisa salvar os dados selecionados ou gerados e de descritivo ou outros dados associados, por identificação, gestão e subsequente reuso.	Metadados; Ferramentas de gestão de referência; Bancos de dados; Repositórios de gestão de dados: políticas e práticas.
5. Usando dados.	5.1. Manipulação de dados.	ser capazes de preparar dados para análise, analisá-los de acordo com os resultados buscados e saber usar o necessário Ferramentas.	Conversão de dados; Manipulando análise de dados ferramentas, tanto localmente (Excel, R, SPSS, Stata ou semelhante) e on-line.

(conclusão)

<i>Framework 02</i>		Competência	Conteúdo
5. Usando dados.	5.2. Produzindo elementos para síntese de dados.	ser capazes de sintetizar e representam os resultados da análise de dados em maneiras adaptadas à natureza dos dados, sua finalidade e o público alvo da consulta.	Escolhendo a representação de dados adequados (tabelas, gráficos ou similares); Manipulação ferramentas (incorporadas em ferramentas analíticas ou aplicativos como Gapminder, Visual.ly ou Muitos Olhos da IBM).
	5.3. Uso ético de dados.	fazer uso ético de dados, reconhecendo a fonte quando obtida ou formulado por outros, e certificando-se de que os métodos usados são implantados e os resultados interpretados de forma transparente e honesta	Qual é o uso ético dos dados; Como citar fontes de dados.

Fonte: Adaptado de Calzada Prado e Marzal, (2013, p. 130).

Listado no Quadro 3.

Quadro 36 - *Framework 03: objetivos de aprendizagem – Foster (2015)*

(continua)

<i>Framework 03</i>		Competência	Conteúdo	
1. Ciência Aberta	1.1 Pesquisa Reprodutível Aberta	Definir relevância para Reprodutibilidade	Identificar ferramentas da Ciência Aberta para cada etapa do ciclo de vida da pesquisa	
			Definir a relevância das ferramentas da Ciência Aberta para a reprodutibilidade / integridade da pesquisa	
		Justificar a abertura como uma ferramenta de reprodutibilidade	Aplicar conceitos de Ciência Aberta os seus processos de pesquisa diários	
			Discutir o papel da Ciência Aberta no processo de revisão por pares	
			Discutir o papel da Ciência Aberta e da Reprodutibilidade em Inovação e Crescimento Econômico	
	1.2 Dados Abertos de Pesquisa	Grande volume de dados abertos	Defina o conceito Grandes Dados Abertos	
			Identificar serviços com base em grandes dados abertos	
		Definição de dados abertos	Definir dados abertos	
			Demonstrar as vantagens dos dados abertos	
		Periódicos de dados abertos	Identificar Revistas de Dados Abertos existentes	
			Preparar uma publicação para um diário de dados abertos	
		Padrões de dados abertos	Identificar padrões de dados abertos existentes	
			Use identificadores para arquivar e citar dados de pesquisa	
		Uso e reutilização de dados abertos	Compreender os dados vinculados (<i>Linked data</i>)	
			Selecionar e usar licenças (por exemplo, CC) para conjuntos de dados	
			Cumprir com os requisitos do projeto piloto Horizon2020 de dados abertos de pesquisa	
		1.3 Acesso Aberto	Distinguir as opções para Acesso aberto	Comparar os custos / benefícios das vias verde e dourada de publicação
				Escolha uma via de publicação relevante com base no seu contexto
	Cumprir com os requisitos do mandato Horizon2020			
	Reconhecer as vantagens do acesso aberto		Analisar o impacto social do acesso aberto	
Estimar o efeito do acesso aberto na visibilidade e impacto dos resultados da pesquisa				
Identificar ferramentas e infra-estrutura eletrônica para o acesso aberto				

(continua)

Framework 03		Competência	Conteúdo
1. Ciência Aberta	1.3 Acesso Aberto	Reutilizar os recursos acesso aberto existentes	Definir as características de uma publicação em acesso aberto
			Use diferentes portais de pesquisa em acesso aberto
			Interpretar licenças de conteúdo e direitos autorais
2. Gestão de dados de pesquisa (Research Data Management - RDM)	2.1 Planos RDM	Definir planos RDM	Indicar características dos planos RDM
			Avaliar as características de um plano de RDM para sua disciplina / projeto
		Preparar um Plano RDM	Criar um plano de RDM
			Aplicar o Plano RDM à sua disciplina / projeto
	2.2 Políticas RDM	Entender as políticas de RDM	Identificar políticas de RDM
			Cumprir com as políticas de RDM
	2.3 Serviços RDM	Identificar serviços RDM	Identificar os serviços RDM relevantes
			Categorizar Serviços RDM
		Usar serviços RDM	Integrar os serviços RDM em sua disciplina / projeto / fluxo de trabalho
	2.4 Padrões RDM	Descrever os Padrões RDM existentes	Identificar padrões de RDM existentes
			Analisar os requisitos das normas RDM
		Implementar Padrões RDM existentes	Promover os Padrões RDM em sua disciplina / projeto
	2.5 Ferramentas RDM	Identificar ferramentas RDM existentes	Identificar ferramentas RDM existentes
			Comparar ferramentas RDM
		Uso de ferramentas RDM	Selecionar ferramentas RDM relevantes
Avaliar o uso de ferramentas RDM			
3. Questões legais	3.1 Direito de propriedade intelectual	Definição de direitos autorais	Definir direitos autorais em Dados abertos, acesso aberto e conteúdo de Ciência Aberta
			Definir complementaridades e conflitos de IPR-OS
		Autores e direitos do editor	Definir limites do contrato de direitos autorais
			Estender um contrato de direitos autorais
			Retor direitos autorais a uma publicação
	3.2 Leis	Identificar leis locais de direitos autorais	Aplicar o acesso aberto via verde à sua disciplina / projeto / fluxo de trabalho
			Entender as leis de direitos autorais relevantes para sua disciplina
	3.3 Licenças	Escolha licenças apropriadas	Diferenciar o nível de direitos de acesso
Identificar tipos de licenças			
	Aplicar a licença	Aplicar a licença adequadas ao seu conteúdo	
4. Ética	4.1 Ética	Delimitar limites éticos	Definir limitações de dados abertos para dados clínicos, privacidade do paciente, segurança nacional

(conclusão)

<i>Framework 03</i>		Competência	Conteúdo
4. Ética	4.1 Ética	Atuar dentro de limites éticos	Relacionar-se com a pesquisa e inovação responsáveis
			Cumprir como Mandato de Acesso Aberto Horizon2020
			Cumprir com o projeto piloto em dados de pesquisa Horizon2020.

Fonte: Adaptado de Foster (2015, *online*).
Listado no Quadro 3.

Quadro 37 - Framework 04: desenvolvimento profissional de pesquisadores em ensino superior – Vitae (2011)

(continua)

<i>Framework 04</i>	Competência	Conteúdo
1. Base de conhecimento	1.1 Métodos de Pesquisa - conhecimento teórico	Percebe o valor em padrões e métodos / técnicas para coleta de informações / dados e análise; avalia e demonstra utilidade e validade das informações / dados no contexto de um problema / questão.
	1.2 Métodos de Pesquisa - aplicação prática	Cria novos modelos e hipóteses, projetos de pesquisa, técnicas de coleta e análise de dados.
		Educa e orienta os outros na seleção e uso apropriado do design de pesquisa, coleta de informações / dados, gestão de informações / dados, análise e métodos / técnicas
	1.3 Busca de informação	Educa os outros em busca de informações / dados, acessando, avaliando e verificando técnicas.
	1.4 Letramento informacional e gestão	Projeta e executa sistemas para a aquisição e o agrupamento de informações usando a tecnologia da informação apropriadamente (por exemplo, processamento de texto, planilhas, sistemas de simulação, bancos de dados). (C2, C4) * Desenvolve a conscientização sobre segurança de informações / dados e questões de longevidade. Sabe onde obter conselhos de especialistas, ou seja, gerenciadores de informações / dados, arquivistas e bibliotecários.
		Desenvolve a conscientização sobre a criação, organização, validação, compartilhamento, armazenamento e curadoria de informações / dados e os riscos associados. Compreende requisitos legais, éticos e de segurança envolvidos na gestão de informações / dados, especialmente ao longo do tempo. Tem conhecimento do objetivo dos metadados
		Aconselha e educa seus pares, pesquisadores, estudantes e funcionários menos experientes em técnicas de gestão de dados / informações específicas da área de pesquisa / disciplina, segurança de dados, requisitos legais e éticos
Desenvolve novas técnicas para gestão de informações. Acompanha e antecipa tendências no desenho e uso de informações / coleta de dados, análise e preservação		
1.5 Letramento acadêmico e numérico	É matematicamente competente para realizar pesquisas na própria disciplina / área de pesquisa; compreende e aplica quaisquer estatísticas que possam ser usadas na área de disciplina / pesquisa; analisa dados e usa pacotes de computador apropriados	
2. Habilidades Cognitivas	2.1 Analisar	Analisa e avalia criticamente os próprios achados e os dos outros. Valida os conjuntos de dados de outras pessoas
	2.2 Sintetizar	Vê conexões entre pesquisa própria e estudos anteriores. Benefícios da orientação com informações e dados sintetizados e idéias

(conclusão)

<i>Framework 04</i>	Competência	Conteúdo
2. Habilidades Cognitivas	2.3 Avaliar	Aconselha e orienta pesquisadores menos experientes sobre a qualidade, integridade, autenticidade e validade de informações / dados de pesquisa primária e secundária.
		Avalia a qualidade, integridade e autenticidade das informações / dados de pesquisa primária e secundária
3. Qualidades pessoais	3.1 Integridade	Atua com integridade profissional e honestidade, toma cuidado especial no manuseio e disseminação de informações / dados e engajamento com os outros.
	3.2 Requerimentos legais	Tem uma compreensão básica dos requisitos legais em torno da pesquisa, por ex. Lei de Proteção de Dados, Lei de Direitos Autorais
		Compreende as obrigações legais da profissão e pode aconselhar seus pares e pesquisadores menos experientes, especialmente sobre a propriedade dos dados e os requisitos da Lei de Proteção de Dados
	3.3 Propriedade Intelectual e direitos autorais	Tem uma compreensão básica das regras de propriedade de dados conforme se aplicam à própria pesquisa
3.4 Meios de comunicação	Está confiante em interações face a face. Utiliza tecnologias de comunicação interativa para networking, compartilhamento de informações / dados e promoção da presença de pesquisa	

**Fonte: Adaptado de Vitae (2011).
Listado no Quadro 3.**

Quadro 38 - *Framework 05: data information literacy project*– Carlson e Johnston (2015).

(continua)

<i>Framework 05</i>	Competência	Conteúdo
1. Letramento Contextual	1.1 Práticas Culturais	Reconhece as práticas, valores e normas do campo escolhido, disciplina ou subdisciplina no que se refere a gerenciar, compartilhar, curar e preservar dados. Reconhece padrões relevantes de dados de campo (por exemplo, metadados, qualidade, formatação) e compreende como esses padrões são aplicados
	1.2 Conversão de Dados e Interoperabilidade	É proficiente na migração de dados de um formato para outro. Compreende os riscos e a possível perda ou corrupção de informações causadas por mudando formatos de dados. Entende os benefícios de disponibilizar dados em formatos padrão para facilitar uso.
	1.3 Curadoria e reutilização de dados	Reconhece que os dados podem ter valor além do propósito original, como validar pesquisa, ou para uso por outras pessoas. É capaz de distinguir quais elementos de um conjunto de dados é provável que tenha valor futuro para si e para os outros. Entende que a curadoria de dados é um esforço complexo, muitas vezes dispendioso, não obstante vital para a pesquisa eletrônica conduzida pela comunidade. Reconhece que os dados devem estar preparados para a sua eventual curadoria na sua criação e ao longo do seu ciclo de vida. Articula o planejamento e as atividades necessárias para permitir a curadoria de dados, tanto geralmente e dentro da prática local. Entende como citar dados, bem como como tornar dados citáveis.
	1.4 Gestão e Organização de Dados	Entende o ciclo de vida dos dados, desenvolve planos de gestão de dados e mantém rastrear a relação de subconjuntos ou dados processados com os conjuntos de dados originais. Cria procedimentos operacionais padrão para gestão e documentação de dados.
	1.5 Preservação de Dados	Reconhece os benefícios e custos da preservação de dados. Entende a tecnologia, os recursos e os componentes organizacionais de preservando dados. Utiliza as melhores práticas na preparação de dados para sua eventual preservação durante sua ciclo de vida ativo. Articula o potencial valor a longo prazo dos próprios dados para si ou para os outros e é capaz para determinar um período de tempo de preservação adequado. Compreende a necessidade de desenvolver políticas de preservação e é capaz de identificar elementos centrais de tais políticas.
	1.6 Processamento e Análise de Dados	Está familiarizado com as ferramentas e técnicas básicas de processamento e análise de dados do disciplina ou área de pesquisa. Entende o efeito que essas ferramentas podem ter sobre os dados. Utiliza ferramentas de gestão de fluxo de trabalho apropriadas para automatizar a análise repetitiva de dados.

(conclusão)

<i>Framework 04</i>	Competência	Conteúdo
1. Letramento Contextual	1.7 Qualidade de Dados e Documentação	Reconhece, documenta e resolve qualquer artefato aparente, incompletude ou corrupção de dados. Utiliza metadados para facilitar a compreensão de possíveis problemas como conjuntos de dados. Documenta dados de forma suficiente para permitir a reprodução de resultados de pesquisa e dados por outros. Rastreia a proveniência de dados e delinea claramente e denota versões de um conjunto de dados.
	1.8 Visualização de dados e representação	Proficientemente usa ferramentas básicas de visualização de disciplina. Evita representações enganosas ou ambíguas ao apresentar dados em tabelas, gráficos e diagramas. Escolhe o tipo apropriado de visualização, como mapas, gráficos, animações, ou vídeos, com base na compreensão do motivo / propósito da visualizando ou exibindo dados.
	1.9 Bancos de dados e formatos de dados	Entende o conceito de bancos de dados relacionais e como consultá-los em bancos de dados. Familiariza-se com formatos e tipos de dados padrão para disciplina. Entende quais formatos e tipos de dados são apropriados para diferentes questões de pesquisas.
	1.10 Descoberta e Aquisição de Dados	Localiza e utiliza repositórios de dados disciplinares. Avalia a qualidade dos dados disponíveis de fontes externas. Identifica não apenas fontes externas de dados apropriadas, mas também importa dados e converte quando necessário para que possa ser usado localmente.
	1.11 Ética e Atribuição de autoria	Desenvolve uma compreensão da propriedade intelectual, privacidade e questões de confidencialidade, eo <i>ethos</i> da disciplina quando se trata de compartilhar e administrando dados. Reconhece os dados de fontes externas . adequadamente. Evita representações enganosas ou ambíguas ao apresentar dados.
	1.12 Metadados e Descrição de Dados	Compreende a lógica dos metadados e anota e descreve proficientemente dados para que possa ser entendido e usado por si e pelos outros. Desenvolve a capacidade de ler e interpretar metadados de disciplinas e fontes externas. Compreende a estrutura e o propósito das ontologias para facilitar um melhor compartilhamento de dados.

**Fonte: Adaptado de Carlson e Johnston (2015).
Listado no Quadro 3.**

Quadro 39 - Framework 06: University of Minnesota Data management course – Jeffryes; Johnston (2013).

(continua)

<i>Framework 06</i>	Competência	Conteúdo
1. Introdução a dados gestão	Conceito de dados gestão usando um exemplo do acadêmico disciplina	Descrever os benefícios dos dados gestão, a fim de explicitamente compreender os benefícios de participando da corrida
2. Dados a serem gerenciados	definir quais informações serão ser gerenciado, documente processo de coleta de dados, e criar um plano para armazenar backup e seguramente abrigam estes dados.	<ol style="list-style-type: none"> 1. Crie um inventário de dados para seu projeto de pesquisa (dados, arquivos de projetos, documentação, etc.) para não ignorar quaisquer aspectos do seu DMP. 2. Escreva um plano de backup e armazenamento para evitar a perda potencial de dados.
3. Organização e documentação métodos	planejar como organizar sua dados, rastrear versões, criar metadados e dados do documento coleção para reutilização.	<ol style="list-style-type: none"> 1. Planeje uma estrutura organizacional para seus dados usando um sistema de nomes de arquivos e estrutura de diretórios que é bem-documentada interoperável com outros conjuntos de dados, a fim de diminuir problemas de versionamento e dados duplicação. 2. Articule um plano para coletar compartilhar os pontos de dados suplementares de sua pesquisa, a fim de ajudar outros pesquisadores em fazer sentido de seus dados. 3. Preencha um esquema de metadados para os seus dados, a fim de práticas de metadados ideais do modelo
4. Acesso a dados e propriedade	Nesta seção nós iremos ilustrar alguns dos propriedade intelectual e preocupações de acesso que enfrentar ao compartilhar seus dados com os outros.	<ol style="list-style-type: none"> 1. Cite as partes interessadas de seus dados, a fim de compreender propriedade intelectual potencial e preocupações de propriedade com a liberação seus dados para um público mais amplo. 2. Relate possíveis preocupações de acesso com seus dados, a fim de planejar controles de acesso apropriados. 3. Identifique os possíveis controles de acesso a fim de proteger seus dados liberação.

(conclusão)

<i>Framework 06</i>	Competência	Conteúdo
5. Compartilhamento de dados e reutilização	benefícios do compartilhamento de dados e potencial para reutilização apresentar aos alunos o conceito de publicação de dados e citação	<ol style="list-style-type: none"> 1. Cite o público para quem os dados serão compartilhados para personalizar a documentação e formato para possível reutilização. 2. Explique uma abordagem que eles usarão para compartilhar os dados, a fim de incutir melhores práticas para seus dados futuros compartilhando. 3. Cite seus dados de maneira apropriada formato estruturado de acordo com padrões emergentes, a fim de prepará-los para reutilizar dados eticamente no futuro.
6. Técnicas de preservação	Introduzir as técnicas de conservação e curadoria utilizadas por profissionais da informação quem gerencia digital informações para longo prazo	<ol style="list-style-type: none"> 1. Explique a vida útil do potencial usar para seus dados, a fim de reconhecer o valor a longo prazo de seus dados. 2. Identifique a preservação relevante amigável formato de arquivo para sua pesquisa dados, a fim de garantir a longo prazo acesso à sua informação digital.
7. Complete seu DMP	Este módulo final ajudará na criação de um plano final de gestão de dados	<ol style="list-style-type: none"> 1. Mapeie um plano de implementação a fim de pôr em prática imediata use o conteúdo dos módulos anteriores. 2. Identifique os componentes de um plano de gestão, a fim de repetir o processo com pesquisas futuras atividades.

**Fonte: Adaptado de Jeffryes; Johnston (2013).
Listado no Quadro 3.**

APÊNDICE H – Unidades de registro de *frameworks*

Quadro 40 - Levantamento de unidades de registro de *frameworks*

(continua)

UNIDADES DE REGISTRO	FRAMEWORK	UNIDADES DE REGISTRO	FRAMEWORK
Academic literacy and numeracy	F04	Innovation	F04
Access	F02	Intellectual Property Rights	F03
Access	F02	Interoperability	F01
Acquisition	F02	Interpret	F02
Acquisition	F01	Interpretation	F02
Acquisition of Data	F01	Introduction to Data Management	F06
Analysing	F04	Introduction to Databases	F01
Analysis	F02	IPR and copyright	F04
Analysis	F01	Legal and ethical	F02
Apply your licence	F03	Legal requirements	F04
Archiving practices	F02	Licences	F03
Argument construction	F04	Locate	F02
Attribution and co-authorship	F04	Long-term access	F01
Authors & Publisher rights	F03	Manage	F02
Backup	F02	Manage	F02
Best practices for data retention	F01	Managing	F02
Collaboration	F04	Managing	F01
Collecting	F02	Managing data	F02
Collecting	F01	Managing data in aggregation	F02
Completing your DMP	F03	Manipulation	F01
Conversion	F01	Manipulation	F01
Critical thinking	F04	Metadata	F02
Critically assess	F02	Metadata	F01
Cultures of Practice	F01	Metadata	F01
Curation	F01	Open	F03
Customize	F02	Open Access	F03
Data Access	F06	Open Data Journals	F03
Data Analysis	F01	Open Data Standards	F03
Data and users	F02	Open Data use and reuse	F03
Data collections	F02	Open Research Data	F03
Data Conversion	F01	Open Science Evaluation	F03
Data curation	F02	Open Science Guidelines	F03
Data Curation	F01	Open Science policies	F03
Data Formats	F01	Open Science Tools	F03
Data integration	F01	Organization	F01
Data management	F02	Organization	F06
Data management	F02	Organizational planning	F02
Data Management	F01	OS Policies	F03
Data presentation	F02	OS Tools	F03
Data Preservation	F01	Overview of research data management	F02
Data Preservation Techniques	F06	Ownership	F06
Data sets description	F02	Plan for archiving and preservation of data	F02
Data sharing	F02	Policy	F04
Data sharing	F02	Problem solving	F04
Data Sharing	F06	Processing	F02
Data sharing and citation	F01	Processing	F01

(conclusão)

UNIDADES DE REGISTRO	FRAMEWORK	UNIDADES DE REGISTRO	FRAMEWORK
Data storage	F02	Publication	F04
Data To be Managed	F06	Quality Assurance	F01
Data Visualization	F01	RDM Plans	F03
Data: types, stages and formats	F02	RDM Standards	F03
Databases	F02	Reading, interpreting and evaluating data	F02
Define the concept of Open Science	F03	Reproducible Research	F03
Directory structures and naming conventions	F01	Research methods – practical application	F04
Discovery	F01	Research methods – theoretical knowledge	F04
Discovery	F01	Research strategy	F04
Documentation and metadata	F01	Re-sharing	F01
Documentation Methods	F06	Respect and confidentiality	F04
Enabling technologies	F02	Reuse	F06
Ethically use data	F02	Re-use	F01
Ethics	F03	Reuse existing OA resources	F03
Ethics, including citation of data	F01	Re-use policies	F02
Ethics, principles and sustainability	F04	Risk management	F04
Evaluate data	F02	Science data life cycle	F02
Evaluating	F02	Scientific data and its description	F01
Evaluating	F01	Security	F02
Evaluating	F04	Security and backups	F01
Evaluation	F02	Society and culture	F04
Extract	F02	Synthesising	F04
Finding and/or obtaining data	F02	Types of data	F02
Fundamentals of science data	F02	Understand	F02
Good file formats	F01	Understanding data	F02
Handle	F02	Use	F02
Handling	F02	Using	F02
Income and funding generation	F04	Using data	F02
Information literacy and management	F04	Using data	F01
Information seeking	F04	Using data ethically	F02
Infrastructure and resources	F04	Visualization	F01

Fonte: Adaptado dos *frameworks* de Calzada Prado e Marzal (2013), Carlson *et al.* (2011), Carlson e Johnston (2015), Foster (2015), Jeffreyes e Johnston (2013) e Vitae (2011), a partir de análises do NVivo (2018).

Listado no Quadro 17.

APÊNDICE I – Unidades de contexto de *frameworks*

Quadro 41 - Levantamento de unidades de contextos de *frameworks*

(continua)

UNIDADES DE CONTEXTO	FRAMEWORK
Ability to access data sources appropriate to the information needed	F02
Ability to apply results to learning, decisionmaking or problem-solving	F02
Ability to critically assess data and their sources	F02
Ability to determine and use suitable research methods	F02
Ability to determine when data are needed	F02
Ability to handle and analyze data	F02
Ability to identify the context in which data are produced and reused (data lifecycle	F02
Ability to plan, organize and self-assess throughout the process	F02
Ability to present quantitative information (specific data, tables, graphs, in reports and similar)	F02
Ability to recognize source data value, types and formats	F02
Analyse existing OS Guidelines	F03
Analyses policies and understands the wider contexts in which they are situated.	F04
Apply lessons learnt for your discipline/project	F03
Apply licence and apply to your content	F03
Apply the RDM Plan to your discipline/project	F03
Appraise the characteristics of a RDM plan to your discipline/project	F03
Appreciates the value of a range of standards and methods/techniques for information/data collection and analysis; assesses and demonstrates usefulness and validity of information/data in the context of a specific problem/question.	F04
Appropriate selection and use of research design, information/data collection, information/data management, analysis and methods/techniques.	F04
Appropriately acknowledges data from external sources.	F01
Articulate a plan to collect and share the supplementary data points of their research	F06
Articulate what they will get out of this program in order to reinforce the learning outcomes of the curriculum	F06
Articulates the planning and actions needed to enable data curation	F01
Assesses the quality, integrity and authenticity of primary and secondary research information/data.	F04
Assesses the reliability, reputation, currency, authority and relevance of sources.	F04
Assist other researchers in making sense of their data.	F06
Avoids misleading or ambiguous representations when presenting data.	F01
Aware of how own research aligns with the research strategy of the institution and strategic focus of the discipline/research area.	F04
Aware of risks in virtual environments and when using interactive communication technologies.	F04
Aware of the value of working collaboratively to benefit research and for maximising the potential for impact.	F04
Becomes familiar with the basic analysis tools of their discipline.	F01
Becomes proficient in migrating data from one format to another	F01
Choose a OS Guideline that applies to your discipline/project	F03
Choose appropriate Licences	F03
Citation practice	F04
Cite their data in a properly structured format in accordance with emerging standards	F06
Competency: learners need to be aware of the role of data in society, how they are generated and by whom, and their possible applications, as well as the implications of their use.	F02

(continua)

UNIDADES DE CONTEXTO	FRAMEWORK
Content: Data evaluation criteria (including authorship, method of obtaining and analyzing data, comparability, inference and data summaries).	F02
Contents: Data producers and consumers; Data lifecycle; Data applications: their impact on science and society; Copyright and licenses influencing data reuse.	F02
Continues to develop academic literacy abilities within wider contexts; understands the literacy requirements for different communication media.	F04
Co-produces research outputs with supervisors/research leaders.	F04
craft workable strategies to access various socioeconomic data;	F02
Create a data inventory for their research project (data, project files, documentation, etc.)	F06
Create a RDM Plan	F03
Creates standard operating procedures for data management and documentation	F01
Critically analyses and evaluates own findings and those of others.	F04
Critically synthesises new and complex information from diverse sources.**	F04
Data and metadata collection and management. Competency: learners need to be aware of the need to save the data selected or generated and of descriptive or other data associated therewith, for due identification, management and subsequent reuse. Content: Metadata; Reference management tools; Databases; Data management repositories: policies and practices.	F02
Data handling. Competency: learners need to be able to prepare data for analysis, analyze them in keeping with the results sought and know how to use the necessary tools. Contents: Data conversion; Handling data analysis tools, both locally (Excel, R, SPSS, Stata or similar) and on-line.	F02
Data sources. Competency: learners need to be aware of the possible data sources, be able to evaluate them and select the ones most relevant to an informational need or a given problem. Contents: Data sources; Criteria for assessing data sources.	F02
Decrease versioning issues and data duplication.	F06
Define copyright in Open Data, Open Access, Open Science content	F03
Define IPR-OS complementarities & conflicts	F03
Define relevance of OS tools to Reproducibility/Integrity of Research	F03
Define relevance of OS tools to Reproducibility/Integrity of Research	F03
Define relevance to Reproducibility	F03
Demonstrates awareness of issues relating to the rights of other researchers, of research subjects, and of others who may be affected by the research	F04
Describe existing RDM Standards	F03
Describe the benefits of data management in order to explicitly understand the benefits of participating in the course	F06
describe the differences between socioeconomic and scientific data	F02
describe what socioeconomic data are about	F02
Designs and executes systems for the acquisition and collation of information using information technology appropriately (e.g. word processing, spreadsheets, simulation systems, databases).	F04
Develops an understanding of intellectual property, privacy and confidentiality issues, and the ethos of their discipline when it comes to sharing data.	F01
Develops awareness of information/data security and longevity issues	F04
Develops awareness of the creation, organisation, validation, sharing, storing and curation of information/data and the associated risks.	F04
Develops awareness of the impact of research on wider society and of the impact of society, the environment and culture on research.	F04
Develops data management plans	F01
Develops the ability to read and interpret metadata from external disciplinary sources	F01
Differentiate level of access rights	F03

(continua)

UNIDADES DE CONTEXTO	FRAMEWORK
Documents and evaluates research processes, using statistics where appropriate.	F04
Ensure long-term access to their digital information.	F06
Estimate the effect of OA on visibility and impact of research results	F03
Ethical use of data. Competency: learners need to make ethical use of data, acknowledging the source when obtained or formulated by others, and making sure that used methods are deployed and results interpreted transparently and honestly. Contents: What is the ethical use of data; How to cite data sources.	F02
Evaluate data quality on the basis of reliability and authority.	F02
Evaluate existing Open Science Tools	F03
Evaluating data. Competency: learners need to be able to evaluate data critically.	F02
Explain an approach they will use to share the data	F06
Explain the lifespan of potential use for their data	F06
Fill out a metadata schema example for their data	F06
Formulates and applies solutions to a range of research problems and effectively analyses and interprets research results	F04
Has basic understanding of data ownership rules as they apply to own research.	F04
Has basic understanding of legal requirements surrounding research	F04
Has broad awareness and knowledge of key relevant funding sources and grant application procedures.	F04
Has knowledge of purpose of metadata.	F04
Has sufficient understanding of copyright, (B2)* IPR, licensing	F04
Identifies and accesses appropriate bibliographical resources, archives and other sources of relevant information (C3)* including web-based resources, primary sources and repositories.	F04
Identify and choose Open Science Tools for your discipline/project	F03
Identify existing Open Data Journals	F03
Identify existing Open Data Standards	F03
Identify local copyright laws	F03
Identify OS tools for each step of the Research Lifecycle	F03
Identify potential access controls	F06
Identify relevant RDM Services	F03
Identify the components of a data management plan	F06
Identify the Guidelines requisites	F03
Identify the relevant preservationfriendly file format for their research data	F06
Identify the suit of Altmetrics for future Research Evaluation	F03
Identify tools and e-infrastructure for OA	F03
Implement existing RDM Standards	F03
instill best practices for their future data sharing.	F06
Integrate RDM services in your discipline/project/workflow	F03
Interpret content licences and copyright	F03
Is IT literate and competent in using information and digital technology.	F04
Is mathematically competent to undertake research in own discipline/research area; understands and applies any statistics that may be used in the discipline/research area; analyses data and uses appropriate computer packages.	F04
Justifies the principles and experimental techniques used in own research	F04
keeps track of the relation of subsets or processed data to the original data sets	F01
Know how to convert one data format into another and be able to establish data access levels.	F02
Knowing how to select and synthesize data and combine them with other information sources and prior knowledge	F02
Knows immediate academic system/work environment, departmental or faculty.	F04
Knows where to obtain expert advice, i.e. information/data managers, archivists and librarians.	F04

(continua)

UNIDADES DE CONTEXTO	FRAMEWORK
Locates and utilizes disciplinary data repositories.	F01
Makes efficient use of available resources.	F04
Makes sound and realistic judgements based on evidence.	F04
Manages the deposit of research outputs, open and wider access, and the Creative Commons license.	F04
Map out an implementation plan in order to poise them for immediate use the previous modules content.	F06
May engage in inter-disciplinary research	F04
Name the audience for whom the data will be shared	F06
Name the stakeholders of their data	F06
Not only identifies appropriate data sources, but also imports data and converts it when necessary, so it can be used by downstream processing tools	F01
Obtaining data. Competency: learners need to be able to detect when a given problem or need cannot be (totally or partially) solved with the existing data and, as appropriate, undertake research to obtain new data. Contents: Main research methods for obtaining original data.	F02
Operate within ethical limits	F03
organize your data, track versions, create metadata and document data	F06
Plan an organizational structure for their data using a file naming system and directory structure that is well documented and interoperable	F06
Plan for the appropriate access controls.	F06
Prepare a publication for an Open Data Journal	F03
Prepare them to ethically reuse data in the future.	F06
Produces some publishable material in print, electronic or other format.	F04
Producing elements for data synthesis. Competency: learners need to be able to synthesize and represent the results of data analysis in ways suited to the nature of the data, their purpose and the audience targeted in the inquiry. Contents: Choosing suitable data representation methods (tables, graphs or similar); Handling tools (built into analytical tools or stand-alone applications such as Gapminder, Visual.ly or IBM's Many Eyes).	F02
Proficiently uses basic visualization tools of discipline.	F01
Promote outputs of OS projects	F03
Promote RDM Standards in your discipline/project	F03
Provides some evidence in support of ideas	F04
Reading and interpreting data. Competency: learners need to be aware of the various forms in which data can be presented (written, numerical or graphic), and their respective conventions, and be able to interpret them. Contents: Ways to present and represent data.	F02
Recognises the importance of bibliometrics and citations.	F04
Recognize the long-term value of their data.	F06
Recognizes and resolves any apparent artifacts, incompleteness, or corruption of data sets.	F01
Recognizes relevant data standards of his/her field (metadata, quality, formatting, etc.) and understands how these standards are applied.	F01
Recognizes that data may have value beyond their original purpose, to validate research or for use by others	F01
Recognizes that data must be prepared for its eventual curation at its creation and throughout its lifecycle	F01
Recognizes the benefits and costs of data preservation	F01
Recognizes the practices, values, and norms of his/her chosen field, discipline or sub-discipline as they relate to managing, sharing, curating and preserving data	F01
Repeat the process with future research activities.	F06
Report potential access	F06
Sees connections between own research and previous studies. Benefits from guidance with synthesising information/data and ideas.	F04

(continua)

UNIDADES DE CONTEXTO	FRAMEWORK
Select & Use licences (e.g. CC) for datasets	F03
students must be aware of data repository policies and purposes	F02
Students must be aware of the various types of data, their formats and data collecting methods and policies.	F02
Students must know what metadata are and how to identify standards and practices applicable to a project.	F02
Students must understand the ethical and legal implications of data use in terms of intellectual property, privacy and citations.	F02
Students must understand the importance of these matters and the best practices against which to benchmark a project.	F02
students must understand the meaning of open science and open data, be aware of the conditions for data reuse	F02
Students must understand what research data are, their life cycle, the need to manage them and the impact of data management on the success of a project.	F02
the ability to analyze and synthesize information to support arguments or generate new ideas	F02
the ability to assess critically both the sources and the ideas expressed therein	F02
the ability to define precisely the informational need	F02
the ability to document the sources used	F02
the ability to locate information sources suited to that need	F02
the ability to manage the information selected	F02
the ability to produce accurate bibliographic citations for data tables	F02
the ability to read an article containing a graphical representation of data and discuss the table in relation to the article content	F02
the ability to read and critically evaluate simple 2x2 or 3-way tables	F02
the ability to record or communicate the results in an ethical manner	F02
the ability to use American FactFinder to create a table, which they could describe and cite correctly;	F02
Understand copyright laws relevant to your discipline	F03
Understand of linked data	F03
Understand RDM Policies	F03
understand the potential intellectual property and ownership concerns with releasing their data to a broader audience.	F06
Understands analytical or statistical procedures in related disciplines/research areas and continues to develop mathematical ability.	F04
Understands and applies the relevant codes of conduct and guidelines for the ethical conduct of research; seeks advice from supervisor.	F04
Understands concept of attribution (B2)* and applies it consistently and fairly to appropriately recognise contributions and co-authorships. Seeks advice on local codes of conduct.	F04
Understands legal, ethical and security requirements involved in information/data management, especially over time.	F04
Understands relevant research methodologies and techniques and their appropriate application within own research area	F04
Understands that curating data is a complex, often costly endeavor that is nonetheless vital to community-driven e-research	F01
Understands the advantages of different types of visualization, for example, maps, graphs, animations, or videos, when displaying data.	F01
Understands the benefits of making data available in standard formats to facilitate downstream use	F01
Understands the concept of relational databases, how to query those databases, and becomes familiar with standard data formats and types for their discipline	F01

UNIDADES DE CONTEXTO	(conclusão) FRAMEWORK
Understands the legal obligations of the profession and can advise peers and less experienced researchers, especially on ownership of data and the requirements of the Data Protection Act.	F04
Understands the lifecycle of data	F01
Understands the rationale for metadata and proficiently annotates and describes data so it can be understood and used by themselves	F01
Understands the risks and potential loss or corruption of information caused by changing data formats	F01
Understands the structure and purpose of ontologies in facilitating better sharing of data	F01
Understands the technology, resource and organizational components of preserving data	F01
Understands the value of open access of research outputs to researchers and the wider society.	F04
Understands which formats and data types are appropriate for different research questions.	F01
Use different OA search portals	F03
Use Identifiers for archiving & citing research data	F03
Use Identifiers for archiving & citing research data	F03
Uses a range of research methods linked to study area; documents own activity.	F04
Uses appropriate workflow management tools to automate repetitive analysis of data.	F01
Utilizes best practices in preservation appropriate to the value and reproducibility of data.	F01
Utilizes metadata to facilitate understanding of potential problems with data sets.	F01
Validates datasets of others.	F04
What is data? Competency: learners need to know what is meant by data and be aware of the various possible types of data. Contents: Data definition; Types of data (depending on origin, format, usage license and so on).	F02
Within own research respects the right of participants to confidentiality and anonymity.	F04
Write a backup and storage plan in order to avoid potential loss of data	F06

Fonte: Adaptado dos *frameworks* de Calzada Prado e Marzal (2013), Carlson *et al.* (2011), Carlson e Johnston (2015), Foster (2015), Jeffreyes e Johnston (2013) e Vitae (2011), a partir de análises do NVivo (2018).

Listado no Quadro 17.

APÊNDICE J – Agrupamento e formação das categorias de análise

Quadro 42 - Agrupamento das categorias de análise dos *frameworks*

(continua)

CATEGORIAS DE ANÁLISE	UNIDADES DE REGISTRO	CATEGORIAS DE ANÁLISE	UNIDADES DE REGISTRO
Access	Access	Archiving	Plan for archiving and preservation of data
	Access		Curation
	Critically assess	Curation	Conversion
	Data Access		Customize
	Discovery		Data Conversion
	Discovery		Data curation
	Information seeking		Data Curation
	Locate		Data Formats
Acquisition	Acquisition		Data integration
	Acquisition		Data: types, stages and formats
	Acquisition of Data		Good file formats
	Collecting		Managing data in aggregation
	Collecting		Manipulation
	Extract		Manipulation
	Finding and/or obtaining data		Processing
			Processing
Analysis	Analysing	Synthesising	
	Analysis	Types of data	
	Analysis	Data management	Completing your DMP
	Argument construction		Data management
	Critical thinking		Data management
	Data Analysis		Data Management
	Evaluate data		Data To be Managed
	Evaluating		Directory structures and naming conventions
	Evaluating		Handle
	Evaluating		Handling
	Evaluation		Information literacy and management
	Interpret		Introduction to Data Management
	Interpretation		Manage
	Problem solving		Manage
	Reading, interpreting and evaluating data		Managing
	Managing		
Archiving	Archiving practices	Managing data	
	Best practices for data retention	Organization	
	Data presentation	Organization	
	Data Preservation	Organizational planning	
	Data Preservation Techniques		
	Long-term access		

(continua)

CATEGORIAS DE ANÁLISE	UNIDADES DE REGISTRO	CATEGORIAS DE ANÁLISE	UNIDADES DE REGISTRO
Data management	Overview of research data management	Policy	Income and funding generation
	RDM Plans		Infrastructure and resources
	RDM Standards		Innovation
	Science data life cycle		Policy
Documentation	Documentation Methods		Respect and confidentiality
	Data sets description		Risk management
	Documentation and metadata		Society and culture
	scientific data and its description	Quality Assurance	
Licences	Apply your licence	Quality	Research methods – practical application
	Attribution and co-authorship		Research methods – theoretical knowledge
	Authors & Publisher rights		Research strategy
	Ethically use data		Reuse
	Ethics	Reuse	Collaboration
	Ethics, including citation of data		Legal requirements
	Ethics, principles and sustainability		Reproducible Research
	Intellectual Property Rights		Re-use
	IPR and copyright		Reuse existing OA resources
	Legal and ethical		Re-use policies
	Licences		Sharing
	Ownership	Data sharing	
	Using data ethically	Data Sharing	
Documentation and metadata	Data sharing and citation		
Metadata	Metadata	Publication	
	Metadata	Re-sharing	
	Metadata	Storage	Backup
	Metadata		Data storage
Metadata	Security		
Open Science	Define the concept of Open Science	Security and backups	
	Open	Using	Data and users
	Open Access		Data collections
	Open Data Journals		Databases
	Open Data Standards		Enabling technologies
	Open Data use and reuse		Fundamentals of science data
	Open Research Data		Good file formats
	Open Science Evaluation		Interoperability
	Open Science Guidelines		Introduction to Databases
	Open Science policies		Understand
	Open Science Tools		Understanding data
	Policy		Academic literacy and numeracy
Cultures of Practice			

(conclusão)

CATEGORIAS DE ANÁLISE	UNIDADES DE REGISTRO
Using	Using
	Using data
	Using data
Visualization	Data Visualization
	Visualization

Fonte: A autoria Própria a partir de análises do NVivo, (2018).

APÊNDICE K – Resultados dos *frameworks* combinados

Quadro 43 - Categorias de análise, unidades de registro e contexto para o reúso de dados segundo *frameworks* combinados

(continua)

Categoria de Contexto	Categoria de Análise	Unidades de Registro	Unidades de Contexto
Frameworks de letramento em dados	Access	Access data sources	Locate information sources suited to that need; Ability to access data sources appropriate to the information needed (F02). Be aware of the possible data sources, be able to evaluate them and select the ones most relevant to an informational need or a given problem. (F02). Use different OA search portals (F03). Differentiate level of access rights (F04). Assesses the reliability, reputation, currency, authority and relevance of sources.(F04)
		Discovery	Locates and utilizes disciplinary data repositories. (F01). Define precisely the informational need (F02). Identifies and accesses appropriate bibliographical resources, archives and other sources of relevant information (C3)* including web-based resources, primary sources and repositories. (F04)
		Access data	Identify potential access controls. Plan for the appropriate access controls. (F06). Report potential access (F06).
	Analysis	Reading	Various forms in which data can be presented (written, numerical or graphic). (F02); know what is meant by data and be aware of the various possible types of data (F02)
		Interpreting/Understanding data	Various forms in which data can be presented (written, numerical or graphic), and their respective conventions, and be able to interpret them. Ways to present and represent data. (F02).
		Evaluating	Evaluate data critically. (F02). Data evaluation criteria (including authorship, method of obtaining and analyzing data, comparability, inference and data summaries). (F02). Analyze them in keeping with the results sought and know how to use the necessary tools. (F02). Uses appropriate workflow management tools to automate repetitive analysis of data. (F01)
		Critical thinking	Critically analyses and evaluates own findings and those of others (F04).
	Acquisition	Obtaining data	Be able to detect when a given problem or need cannot be (totally or partially) solved with the existing data and, as appropriate, undertake research to obtain new data. (F02).
		Collecting Data	Data collecting methods and policie. (F02); Articulate a plan to collect and share the supplementary data points of their research (F06). Executes systems for the acquisition and collation of information using information technology appropriately (e.g. word processing, spreadsheets, simulation systems, databases). (F04)
	Archiving	Plan for archiving	Recognizes the benefits and costs of data preservation (F01). Use Identifiers for archiving (F03).
		Archiving practices	Utilizes best practices in preservation appropriate to the value and reproducibility of data. (F01). Use Identifiers for archiving (F03). Awareness of information/data security and longevity issues (F04). Ensure long-term access to their digital information. (F06).

(continua)

Categoria de Contexto	Categoria de Análise	Unidades de Registro	Unidades de Contexto
<i>Frameworks de letramento em dados</i>	<i>Archiving</i>	<i>Data Preservation Techniques</i>	<i>Understands the technology, resource and organizational components of preserving data (F01). Identify the relevant preservationfriendly file format for their research data (F06)</i>
		<i>Conversion</i>	<i>Becomes proficient in migrating data from one format to another (F01). Know how to convert one data format into another and be able to establish data access levels. (F02)</i>
	<i>Curation</i>	<i>Formats</i>	<i>Understands the risks and potential loss or corruption of information caused by changing data formats (F01); Understands which formats and data types are appropriate for different research questions (F01); Ability to recognize source data value, types and formats (F02).</i>
		<i>Merging</i>	<i>Recognizes and resolves any apparent artifacts, incompleteness, or corruption of data sets (F01);</i>
		<i>Processing</i>	<i>Articulates the planning and actions needed to enable data Curation (F01).</i>
		<i>Synthesize</i>	<i>Synthesize and represent the results of data analysis in ways suited to the nature of the data, their purpose and the audience targeted in the inquiry. Choosing suitable data representation methods (tables, graphs or similar); synthesize information to support arguments or generate new ideas (F02). Critically synthesises new and complex information from diverse sources.(F04)</i>
		<i>Data Management Plan</i>	<i>Identify the components of a data management plan (F06). Develops data management plans (F01).</i>
	<i>Data Management</i>	<i>Directory structures</i>	<i>Plan an organizational structure for their data using a directory structure that is well documented and interoperable (F06)</i>
		<i>Naming conventions</i>	<i>Plan an organizational structure for their data using a file naming system that is well documented and interoperable (F06)</i>
		<i>Organizational planning</i>	<i>Ability to plan, organize and self-assess throughout the process (F02).</i>
		<i>Standards</i>	<i>Research data are, their life cycle, the need to manage them and the impact of data management on the success of a project. (F02). Recognizes relevant data standards of his/her field (metadata, quality, formatting, etc.) and understands how these standards are applied. (F01). Promote RDM Standards in your discipline/project(F03) Understand RDM Policies (F03)</i>
		<i>Documentation</i>	<i>Data sets description</i>
	<i>Documentation Methods</i>		<i>Documents own activity. (F04). Create a data inventory for their research project (data, project files, documentation, etc.) (F06).</i>
	<i>Licences</i>	<i>Apply your licence</i>	<i>Choose appropriate Licences (F04). Select & Use licences (e.g. CC) for datasets (F03). Has basic understanding of data ownership rules as they apply to own research. (F04).</i>

(continua)

Categoria de Contexto	Categoria de Análise	Unidades de Registro	Unidades de Contexto
Frameworks de letramento em dados	Licences	<i>Citation</i>	<i>How to cite data sources (F02). Produce accurate bibliographic citations for data tables (F02). Use Identifiers for citing research data (F03). Cite their data in a properly structured format in accordance with emerging standards (F06).</i>
		<i>Intellectual Property Rights</i>	<i>Understand the ethical and legal implications of data use in terms of intellectual property, privacy and citations (F02). Understand the potential intellectual property and ownership concerns with releasing their data to a broader audience. (F06).</i>
		<i>Licenses types and copyright</i>	<i>Define copyright in Open Data, Open Access, Open Science content (F04). Identify local copyright laws (F03). Interpret content licences and copyright (F03). Understand copyright laws relevant to your discipline (F03).</i>
		<i>Publisher rights</i>	<i>Document the sources used (F02)</i>
		<i>Using data ethically</i>	<i>Acknowledging the source when obtained or formulated by others; record or communicate the results in an ethical manner (F02). Awareness of issues relating to the rights of other researchers, of research subjects, and of others who may be affected by the research (F04). Understands and applies the relevant codes of conduct and guidelines for the ethical conduct of research; seeks advice from supervisor. (F04)</i>
		Metadata	<i>Metadata Description</i>
	<i>Metadata Standards</i>		<i>How to identify standards and practices applicable to a project. (F02). Has knowledge of purpose of metadata. (F04)</i>
	Open Science	<i>Concept of Open Science</i>	<i>Understand the meaning of open science and open data (F02). Understands the value of open access of research outputs to researchers and the wider society.(F04)</i>
		<i>Open Access</i>	<i>Identify tools and e-infrastructure for OA (F03)</i>
		<i>Open Data Journals</i>	<i>Identify existing Open Data Journals (F03). Prepare a publication for an Open Data Journal(F03)</i>
		<i>Open Data Standards</i>	<i>Identify existing Open Data Standards (F03)</i>
	Policy	<i>Open Data use and reuse</i>	<i>Identify OS tools for each step of the Research Lifecycle (F03); Define relevance of OS tools to Reproducibility/Integrity of Research (F03);</i>
		<i>Funding policies</i>	<i>Be aware of the role of data in society, how they are generated and by whom, and their possible applications, as well as the implications of their use. (F02). Has broad awareness and knowledge of key relevant funding sources and grant application procedures. (F04)</i>

(conclusão)

Categoria de Contexto	Categoria de Análise	Unidades de Registro	Unidades de Contexto
<i>Frameworks de letramento em dados</i>	<i>Quality</i>	<i>Quality Assurance</i>	<i>Evaluate data quality on the basis of reliability and authority. (F02). Assesses the quality, integrity and authenticity of primary and secondary research information/data. (F04).</i>
		<i>Research methods – practical application</i>	<i>Ability to determine and use suitable research methods (F02)</i>
		<i>Research methods – theoretical knowledge</i>	<i>Understands relevant research methodologies and techniques and their appropriate application within own research area (F04)</i>
	<i>Reuse</i>	<i>Reproducible Research</i>	<i>Ability to identify the context in which data are produced and reused (data lifecycle) (F02). Recognizes that data may have value beyond their original purpose, to validate research or for use by others (F01). Prepare them to ethically reuse data in the future. (F06)</i>
		<i>Legal requirements</i>	<i>Be aware of the conditions for data reuse (F02). Has basic understanding of legal requirements surrounding research (F04)</i>
		<i>Re-use policies</i>	<i>Be aware of data repository policies and purposes (F02). Explain the lifespan of potential use for their data (F06). Explain an approach they will use to share the data (F06). Decrease versioning issues and data duplication. (F06).</i>
	<i>Sharing</i>	<i>re-sharing</i>	<i>Understands the benefits of making data available in standard formats to facilitate downstream use (F01). Instill best practices for their future data sharing. (F06)</i>
	<i>Storage</i>	<i>Backup</i>	<i>Write a backup and storage plan in order to avoid potential loss of data (F06)</i>
		<i>Security</i>	<i>Understands security requirements involved in information/data management, especially over time (F04). Develops awareness of information/data security and longevity issues (F04).</i>
	<i>Using</i>	<i>Databases</i>	<i>Understands the concept of relational databases, how to query those databases, and becomes familiar with standard data formats and types for their discipline (F01)</i>
		<i>Data collections</i>	<i>Ability to determine when data are needed (F02).</i>
	<i>Visualization</i>	<i>Visualization tools</i>	<i>Ability to present quantitative information (specific data, tables, graphs, in reports and similar) (F02); Avoids misleading or ambiguous representations when presenting data. (F01); Understands the advantages of different types of visualization, for example, maps, graphs, animations, or videos, when displaying data. (F01); Proficiently uses basic visualization tools of discipline.(F01)</i>

Fonte: Adaptado dos *frameworks* de Calzada Prado e Marzal (2013), Carlson *et al.* (2011), Carlson e Johnston (2015), Foster (2015), Jeffreyes e Johnston (2013) e Vítac (2011), a partir de análises do NVivo (2018).

Listados na p. 187, Quadro 19.

APÊNDICE L – Resultados do reúso de dados de Kim (2016 [datasets])

Tabela 21 - Resultados dos conjuntos de dados analisados 1

(continua)

Seção	Pergunta	ID Pergunta	Ciências Sociais		Demais Disciplinas	
			122	2.202	N	%
Metadata	Researchers can easily access metadata.	Metadata1	30	25%	737	33%
	Metadata is available for researchers to share data.	Metadata2	26	21%	817	37%
	Researchers have the metadata necessary to share data.	Metadata3	21	17%	680	31%
Data Repositories	Researchers can easily access data repositories.	Repository1	76	62%	1302	59%
	Data repositories are available for researchers to share data.	Repository2	81	66%	1396	63%
	Researchers have the data repositories necessary to share data.	Repository3	58	48%	1142	52%
Altruism	I am willing to help other researchers by sharing data.	Altruism1	103	84%	1.977	90%
	I would share data so that other researchers can conduct their research more easily.	Altruism2	102	84%	1.975	90%
	I would share data so that other researchers can utilize it for their research.	Altruism3	99	81%	1.920	87%
Perceived Career Benefits	I can earn academic credit such as more citations by sharing data.	Benefit1	46	38%	1089	49%
	Data sharing would enhance my academic recognition.	Benefit2	57	47%	1280	58%
	Data sharing would improve my status in a research community.	Benefit3	62	51%	1392	63%
	Data sharing would be helpful in my academic career.	Benefit4	44	36%	1170	53%
Perceived Career Risks	There is a high probability of losing publication opportunities if I share data.	Risk1	58	48%	1064	48%
	Data sharing may cause my research ideas to be stolen by other researchers.	Risk2	61	50%	1161	53%
	My shared data may be misused or misinterpreted by other researchers.	Risk3	74	61%	1397	63%
	I believe that the overall riskiness of data sharing is high.	Risk4	51	42%	798	36%

(continua)

Pergunta	ID Pergunta	Ciências Sociais		Demais Disciplinas		
		122		2.202		
		N	%	N	%	
Perceived Efforts	Sharing data involves too much time for me (e.g. to organize/annotate).	Effort1	72	59%	1254	57%
	I need to make a significant effort to share data.	Effort2	74	61%	1421	65%
	I would find data sharing difficult to do.	Effort3	59	48%	920	42%
	Overall, data sharing requires a significant amount of time and effort.	Effort4	75	61%	1457	66%
Data Sharing	Sharing data is valuable.	Attitude1	111	91%	2014	91%
	Sharing data is desirable.	Attitude2	108	89%	1943	88%
	Sharing data is pleasant.	Attitude3	44	36%	1009	46%
	Sharing data is interesting.	Attitude4	79	65%	1511	69%
Altruism	I would share data to support open scientific research.	AltruismCommunity1	102	84%	1934	88%
	I would share data to contribute to better scientific research.	AltruismCommunity2	104	85%	2015	92%
	I would share data to help improve the quality of scientific research.	AltruismCommunity3	103	84%	1999	91%
Data Sharing Frequencies	Deposited your data into <u>disciplinary data repositories</u> for every article?	MethodDisciplinaryRepository	3	2%	309	14%
	Deposited your data into <u>institutional data repositories</u> for every article?	MethodInstitutionalRepository	7	6%	233	11%
	Uploaded your data into “ <u>public</u> ” Web spaces for every article?	MethodPublicWeb	8	7%	302	14%
	Provided access to your data by <u>publishing supplement materials</u> for every article?	MethodSupplement	5	4%	527	24%
	Been personally asked to share data for each article?	MethodRequested	5	4%	177	8%
	Responded to the request(s) by providing data via personal communication methods (e.g. email)?	MethodResponded	15	12%	575	26%

(conclusão)

Seção	Pergunta	ID Pergunta	Ciências Sociais		Demais Disciplinas	
			N	%	N	%
			122		2.202	
Resources	Resources that promote data sharing are available to me.	Resource1	30	25%	722	33%
	Specialized instruction concerning data sharing is available to me.	Resource2	17	14%	468	21%
	A specialized person (or group) is available to assist with my data sharing.	Resource3	15	12%	379	17%
	Information technologies are available to support my data sharing.	Resource4	39	32%	781	35%
	My organization is very supportive of my data sharing.	Resource5	20	16%	560	25%

Fonte: Autoria Própria (2019) a partir de análises das respostas de Kim, (2016 [datasets]).

APÊNDICE M – Resultados do reuso de dados de Kim (2017 [datasets])

Tabela 22 - Resultados dos conjuntos de dados analisados 2

Seção	Pergunta	ID Pergunta	Ciências Sociais		Demais Disciplinas	
			176	1.240	N	%
Usefulness	Improves the quality of my research	Usefulness1	93	53%	721	58%
	Enhances the effectiveness of my research	Usefulness2	101	57%	755	61%
	Reduces the time/cost/effort I spend on my research	Usefulness3	96	55%	624	50%
Concerns	I might misinterpret the data	Concern1	66	38%	480	39%
	I might cause infringement	Concern2	30	17%	279	23%
	I might not publish with that data	Concern3	23	13%	241	19%
Efforts	To locate data sets	Effort1	99	56%	836	67%
	To access (or get permission to use) data sets	Effort2	92	52%	728	59%
	To process data sets for a new study	Effort3	100	57%	799	64%
Data Repositories	Data repositories are available for researchers to share data	Repository1	105	60%	702	57%
	Researchers can easily access data repositories to reuse data	Repository2	82	47%	498	40%
Organizational Resources	Resources that promote data reuse are available to me.	OrgResource1	76	43%	454	37%
	Information technologies are available to support my data reuse.	OrgResource2	89	51%	559	45%
	A specialized person is available to assist with my data reuse	OrgResource3	37	21%	186	15%
Intention to Reuse	I am likely to reuse other researchers' data for my future research.	Intention1	95	54%	651	53%
	I intend to reuse other researchers' data for my future research.	Intention2	74	42%	524	42%
	I will try to reuse other researchers' data for my future research.	Intention3	72	41%	548	44%

Fonte: Autoria Própria (2019) a partir de análises das respostas de Kim, (2017 [datasets]).

APÊNDICE N – Resultados do reúso de dados de Berghmans *et al.*, 2017 [*datasets*]

Tabela 23 - Resultados dos conjuntos de dados analisados 3

(continua)

Seção	Pergunta	ID Pergunta	Ciências Sociais		Demais Disciplinas	
			75		1.200	
			N	%	N	%
Categories of research data used	Observational or empirical: e.g. sensor data, survey data, interview transcripts, sample data, neuroimages, ethnographic data, diaries.	Q1a_1	65	87%	648	54%
	Experimental: e.g. gene sequences, chromatograms, toroid magnetic field data.	Q1a_2	2	3%	407	34%
	simulation: e.g. climate models, economic models.	Q1a_3	0	0%	245	20%
	Derived or compiled: e.g. text and data mining, compiled database, 3D models.	Q1a_4	16	21%	278	23%
	Reference or canonical: e.g. gene sequence databanks, chemical structures, spatial data portals.	Q1a_5	3	4%	137	11%
Research data format(s) used	Text: e.g. flat text files.	Q1b_1	36	48%	508	42%
	Numerical: e.g. SPSS, Stata, Excel.	Q1b_2	60	80%	799	67%
	Databases: e.g. MS Access, SQL Server, MySQL, Oracle.	Q1b_3	13	17%	216	18%
	Multimedia: e.g. jpeg, tiff, mpeg, QuickTime, mp3, DICOM.	Q1b_4	11	15%	371	31%
	Models: e.g. 3D, statistical.	Q1b_5	13	17%	320	27%
	Software: e.g. Java, C.	Q1b_6	3	4%	268	22%
	Discipline specific: e.g. FITS in astronomy, CIF in chemistry.	Q1b_7	1	1%	103	9%
	Instrument specific: e.g. proprietary microscope formats such as Olympus Confocal Microscope Data Format, Carl Zeiss Digital Microscopic Image Format (ZVI).	Q1b8	0	0%	156	13%

(continua)

Seção	Pergunta	ID Pergunta	Ciências Sociais		Demais Disciplinas	
			75		1.200	
			N	%	N	%
Research data format(s) to be shared	Text: e.g. flat text files.	Q1c_1	20	27%	352	29%
	Numerical: e.g. SPSS, Stata, Excel.	Q1c_2	53	71%	642	54%
	Databases: e.g. MS Access, SQL Server, MySQL, Oracle.	Q1c_3	7	9%	136	11%
	Multimedia: e.g. jpeg, tiff, mpeg, QuickTime, mp3, DICOM.	Q1c_4	7	9%	229	19%
	Models: e.g. 3D, statistical.	Q1c_5	8	11%	215	18%
	Software: e.g. Java, C.	Q1c_6	3	4%	159	13%
	Discipline specific: e.g. FITS in astronomy, CIF in chemistry.	Q1c_7	0	0%	68	6%
	Instrument specific: e.g. proprietary microscope formats such as Olympus Confocal Microscope Data Format, Carl Zeiss Digital Microscopic Image Format (ZVI).	Q1c_8	0	0%	75	6%
Research data ownership - before publication	Funder	Q1ei_1	22	29%	258	22%
	Institute	Q1ei_2	22	29%	428	36%
	My department	Q1ei_3	19	25%	316	26%
	Colleagues/collaborators	Q1ei_4	26	35%	453	38%
	Myself	Q1ei_5	47	63%	736	61%
	Publisher	Q1ei_6	1	1%	59	5%
	Other, please specify_____	Q1ei_98	4	5%	38	3%
	Don't know [EXCLUSIVE PER COLUMN]	Q1ei_99	2	3%	37	3%
Research data ownership - after publication	Funder	Q1eii_1	26	35%	321	27%
	Institute	Q1eii_2	22	29%	470	39%
	My department	Q1eii_3	22	29%	321	27%
	Colleagues/collaborators	Q1eii_4	23	31%	423	35%
	Myself	Q1eii_5	36	48%	528	44%
	Publisher	Q1eii_6	19	25%	412	34%
	Other, please specify_____	Q1eii_98	5	7%	73	6%
	Don't know [EXCLUSIVE PER COLUMN]	Q1eii_99	5	7%	65	5%
Publishing research data	As an appendix to a peer-reviewed research publication	Q1f_1	12	16%	336	28%
	As a stand-alone peer-reviewed data publication	Q1f_2	11	15%	353	29%
	In a data repositior	Q1f_3	14	19%	153	13%

(continua)

Seção	Pergunta	ID Pergunta	Ciências Sociais		Demais Disciplinas	
			75		1.200	
			N	%	N	%
Attitudes towards research data sharing and reuse	Having access to others' research data benefits/would benefit my own research	Q2a_1	50	67%	870	73%
	I am willing to allow others to access my research data	Q2a_2	48	64%	766	64%
	I have previously shared my research data with others	Q2a_3	50	67%	767	64%
	I encounter obstacles sharing my research data with others	Q2a_5	27	36%	408	34%
	I have received sufficient training in research data sharing	Q2a_8	20	27%	287	24%
	Research data management specialists need to play a role in research data sharing	Q2a_10	41	55%	673	56%
Benefits of publishing research data alongside article	Article more likely to be accepted for publication	Q2c_1	7	9%	234	20%
	Article more likely to be cited	Q2c_2	8	11%	299	25%
	Compliance with funding body mandates	Q2c_3	7	9%	98	8%
	Compliance with journal or publisher requirements	Q2c_4	9	12%	153	13%
	Reproducibility of research	Q2c_5	14	19%	339	28%
	Data reuse	Q2c_6	11	15%	254	21%
	Research aggregation	Q2c_7	11	15%	228	19%
	More possibilities for collaboration	Q2c_8	12	16%	356	30%
	Encourage other researchers to make their data publicly available	Q2c_9	8	11%	285	24%
	Other, please specify	Q2c_98	9	12%	6	1%
Don't know	Q2c_99	0	0%	11	1%	
Research data management and archiving Do you take steps to manage your research data and/or archive it for potential re-use by yourself and/or others?	Yes- Always	Q2e_1	37	49%	541	45%
	Sometimes	Q2e_2	24	32%	551	46%
	Never	Q2e_3	12	16%	11	1%

(continua)

Seção	Pergunta	ID Pergunta	Ciências Sociais		Demais Disciplinas	
			75		1.200	
			N	%	N	%
Whether will use Creative Commons	CC 0 (no rights reserved by the creator)	Q3d_1	1	1%	42	4%
	CC BY Attribution (allows others, even commercially, to distribute and modify your work as long as they credit you for the original creation)	Q3d_2	6	8%	124	10%
	CC BY-NC NonCommercial (allows non-commercial use but others do not have to license derivative work on the same terms)	Q3d_3	7	9%	124	10%
	CC BY-NC-ND Attribution-NonCommercial-NoDerivs (others can download your work and share, as long as they credit you, but cannot modify or use your work commercially)	Q3d_4	13	17%	171	14%
	Other, please specify _____	Q3d_98	2	3%	32	3%
	Don't know	Q3d_99	45	60%	728	61%
	Trust in others' research data. How important were the following factors when deciding to use others' research data that the data... - Important, Somewhat important, Very important	Are from someone I know personally	Q3e_1	23	31%	391
Are from someone I DON'T know personally but who has a good reputation		Q3e_2	26	35%	431	36%
Are from someone at an institute with a good reputation		Q3e_3	27	36%	450	38%
Are from a repository with a good reputation		Q3e_4	27	36%	461	38%
Comply with guidelines around collection and/or formatting		Q3e_5	28	37%	480	40%
Have good documentation and/or metadata		Q3e_6	28	37%	515	43%
Are certified or accredited by a third party		Q3e_7	21	28%	403	34%
Are commonly used by others		Q3e_8	24	32%	435	36%
Have been cited in a publication in the formal literature		Q3e_9	26	35%	487	41%

(conclusão)						
Seção	Pergunta	ID Pergunta	Ciências Sociais		Demais Disciplinas	
			75	1.200	N	%
			N	%	N	%
Citing others' research data	There are no agreed standards for citing research data in my field	Q3f_1	43	57%	554	46%
	There are agreed standards for citing research data in my field, but most researchers follow them	Q3f_2	15	20%	449	37%
	There are agreed standards for citing research data in my field, but most researchers do not follow them	Q3f_3	10	13%	84	7%
	There is no need for to have standards for the citation of research data	Q3f_4	5	7%	79	7%

Fonte: Autoria Própria (2019) a partir de análises das respostas de Berghmans *et al.* (2017 [datasets]).

APÊNDICE O – Estratégia de Busca Textual – Recursos de Aprendizagem Online

(continua)

Nome	Descrição
01. Access	
01.1 Access data sources	"Access data sources" OR "access data repository"
01.2 Discovery	"Discovery data" OR "Discovery dataset"
01.3 Access data	"Access data" OR "Access dataset"
02. Analysis	
02.1 Reading	"Reading data" OR "Read data"
02.2 Interpreting Understanding data	"Interpreting data" OR "Understanding data"
02.3 Evaluating	"Evaluating data"
02.4 Critical thinking	"Critical thinking"
03. Acquisition	
03.1 Obtaining data	"Obtaining data"
03.2 Collecting Data	"Collecting Data"
04. Archiving	
04.1 Plan for archiving	"Plan for archiving" OR "Archive planning" OR "Archive plan" OR "Archiving plan"
04.2 Archiving practices	"Archiving practices" OR "Archiving data"
04.3 Data Preservation Techniques	"Data Preservation Techniques" OR "Preservation practices"
05. Curation	
05.1 Data Conversion	"Data Conversion"
05.2 Data Format	"Data Format" OR "Data Formats"
05.3 Merging	"Data Merging" OR "Data Merge" "Merging Data" OR "Merge Data" OR "Merge Files" OR "Merging Files"
05.4 Data Processing	"Data Processing"
05.5 Data Synthesize	"Data Synthesize" OR "Data Synthesizing" OR "Synthesize Data" OR "Synthesizing Data" OR "Summarizing Data"
06. Data Management	
06.1 Data Management Plan	"Data Management Plan"
06.2 Directory structures	"Directory structures" OR "Directories structures" OR "Directory structure" OR "Files structures" OR "Files structure" OR "File structure"
06.3 Naming conventions	"Naming convention" OR "Naming conventions"
06.4 Organizational planning	"Organizational planning" OR "Data Organizing" OR "Organizing Data" OR "Data organizational planning" OR "Data organization"
06.5 Standards	"Data Standard" OR "Data Standards"
07. Documentation	
07.1 Data sets description	"Data sets description" OR "Datasets description" OR "Dataset description" OR "Data set description" OR "Data description"
07.2 Documentation Methods	"Documentation Methods"

(continua)

Nome	Descrição
08. Licences	
08.1 Apply your licence	"Apply your licence" OR "licence application" OR "choose the licence" OR "choosing licence" OR "licence option"
08.2 Citation	"Data Citation"
08.3 Intellectual Property Rights	"Intellectual Property Rights"
08.4 Licenses types and copyright	License
08.5 Publisher rights	"Publisher rights" OR "Publisher right" OR "Publisher license" OR "Publishing rights" OR "Publishing licenses" OR "Publisher copyright" OR "Publishing copyright"
08.6 Using data ethically	"Using data ethically" OR "Data ethical use" OR "Ethically reuse" OR "Data Ethically use"
09. Metadata	
09.1 Metadata Description	"Metadata Description" OR "Metadata Documentation"
09.2 Metadata Standards	"Metadata Standards"
10. Open Science	
10.1 Concept of Open Science	" Open Science Concept"
10.2 Open Access	"Open Access"
10.3 Open Data Journals	"Open Data Journals"
10.4 Open Data Standards	"Open Data Standards" OR "Open Data pattern" OR "Open Data patterns" OR "Data Standards"
10.5 Open Data use and reuse	"Open Data use and reuse"
11. Policy	
11.1 Funding policies	"Funding policies"
12. Quality	
12.1 Quality Assurance	"Quality Assurance" OR "Data Quality"
12.2 Research methods – practical application	"Research methods"
12.3 Research methods – theoretical knowledge	"theoretical knowledge"
13. Reuse	
13.1 Reproducible Research	"Reproducible Research"
13.2 Legal requirements	"Reuse Legal requirements" AND reuse
13.3 Re-use policies	"Re-use policies" OR "Reuse policies"
14. Sharing	
14.1 Re-sharing	"Re-sharing"
15. Storage	
15.1 Backup	"Backup" OR "security copy"
15.2 Security	"Security data"
16. Using	
16.1 Databases	"Data repository" OR "Data repositories"

(conclusão)

Nome	Descrição
16.2 Data collections	"Data collections"
17. Visualization	
17.1 Visualization tools	"Visualization tool"
18. Support	
18.1 Data Support	"Data support"

Fonte: A autoria própria (2019).

APÊNDICE P – Recorrência de todos os requisitos em cada recurso de aprendizagem *online*

No Quadro 44 é possível verificar a recorrência de cada requisito em cada recurso analisado:

Quadro 44 - Recursos de Aprendizagem Online por Recorrência de Codificação

Requisitos por recursos de aprendizagem - quantidade	A : Curso 01- Bodleian	B : Curso 02- CESSDA	C : Curso 03-DM Course	D : Curso 04- DataONE	E : Curso 05-DC 101	F : Curso 06- Essentials 4 DS	G : Curso 07- FOSTER-OD	H : Curso 08-RD Handbook	I : Curso 09- MANTRA	J : Curso 10- NECDMC	K : Curso 11-RDM- Coursera	TOTAL
1 : 01. Access	19	82	11	28	30	8	22	26	48	61	18	353
2 : 02. Analysis	1	11	0	12	0	5	3	0	15	6	2	55
3 : 03. Acquisition	4	9	1	0	4	22	0	1	5	6	7	59
4 : 04. Archiving	0	18	2	2	16	0	0	0	0	21	0	59
5 : 05. Curation	4	29	1	26	2	39	3	9	13	9	0	135
6 : 06. Data Management	21	56	38	36	64	23	23	10	55	143	39	508
7 : 07. Documentation	0	2	5	2	0	1	1	2	2	6	0	21
8 : 08. Licences	5	119	33	134	91	10	29	6	38	109	46	620
9 : 09. Metadata	6	9	4	18	3	5	7	8	3	34	9	106
10 : 10. Open Science	1	44	5	12	21	28	32	12	6	33	6	200
11 : 11. Policy	0	0	0	0	0	0	0	0	0	0	0	0
12 : 12. Quality	1	7	0	16	6	5	1	7	3	6	5	57
13 : 13. Reuse	0	0	0	0	11	18	0	0	1	6	1	37
14 : 14. Sharing	0	0	0	0	0	0	1	0	0	0	0	1
15 : 15. Storage	8	77	26	36	52	18	6	13	48	160	26	470
16 : 16. Using	20	203	34	50	37	29	20	63	28	54	43	581
17 : 17. Visualization	0	0	0	2	4	0	0	0	0	6	0	12
18 : 18. Support	1	2	1	2	1	113	2	3	1	4	0	130
TOTAL	91	668	161	376	342	324	150	160	266	664	202	3404

Fonte: Autoria própria (2019) com base em nas buscas realizadas nos recursos listados no Quadro 4.

	A : Curso 01- Bodleian	B : Curso 02- CESSDA	C : Curso 03-DM Course	D : Curso 04- DataONE	E : Curso 05-DC 101	F : Curso 06- Essentials 4 DS	G : Curso 07- FOSTER-OD	H : Curso 08-RD Handbook	I : Curso 09- MANTRA	J : Curso 10- NECDMC	K : Curso 11-RDM- Coursera	TOTAL
19 : 09.1 Metadata Description	3	8	0	10	1	0	3	5	0	1	0	31
20 : 09.2 Metadata Standards	3	5	7	12	4	8	4	4	7	43	10	107
21 : 10.1 Concept of Open Science	1	5	0	0	6	1	26	0	1	13	1	54
22 : 10.2 Open Access	0	33	1	4	20	26	25	7	6	28	4	154
23 : 10.3 Open Data Journals	0	2	0	0	0	0	1	3	0	0	0	6
24 : 10.5 Open Data use and reuse	4	9	4	10	11	2	9	2	4	13	2	70
25 : 11.1 Funding policies	2	0	0	0	1	1	0	0	0	2	0	6
26 : 12.1 Quality Assurance	1	7	0	14	4	0	0	6	2	5	5	44
27 : 12.2 Research methods – practical	0	0	0	2	2	5	1	1	1	1	0	13
28 : 13.1 Reproducible Research	0	0	0	0	3	18	0	0	1	0	1	23
29 : 13.2 Legal requirements	0	3	0	0	1	0	0	0	0	1	1	6
30 : 13.3 Re-use policies	0	0	0	0	12	0	0	0	0	24	0	36
31 : 14.1 Re-sharing	0	2	0	0	0	0	1	0	0	13	0	16
32 : 16.1 Databases	8	133	25	18	17	15	15	4	7	33	29	304
33 : 16.2 Data collections	12	74	9	32	20	15	5	59	21	21	15	283
34 : 17.1 Visualization tools	0	0	0	2	4	0	0	0	0	6	0	12
TOTAL	87	586	125	342	307	204	173	143	230	553	177	2927

Fonte: Autoria própria (2019) com uso do NVivo e com base em nas buscas realizadas nos recursos listados no Quadro 4

(conclusão)

Best Practice	Access	Compre-Hension	Discovera-Bility	Interopera-Bility	Processa-Bility	Linka-Bility	Reuse	Trust
Assess dataset coverage							X	X
Gather feedback from data consumers		X					X	X
Make feedback available							X	X
Enrich data by generating new data		X			X		X	X
Provide Complementary Presentations	X	X					X	X
Provide Feedback to the Original Publisher				X			X	X
Follow Licensing Terms							X	X
Cite the Original Publication			X				X	X

Fonte: Adaptado de Lóscio, Burle e Calegari (2017).

ANEXO A – Instrumento de coleta de dados 1

Scientists' Data Sharing Behavior Survey
(KIM, 2016)

(Page 1)

NOTE: In this survey, *Data Sharing* means providing the raw data of your published articles to other researchers outside your research group(s) by making it accessible through data repositories/ public web spaces/ supplementary materials or by sending the data via personal communication methods upon request.

ABOUT YOUR DISCIPLINE

1. Which one of the following best describes your primary subject discipline based on your current research?
(Dropdown Selection Provided) (Discipline)

Please indicate to what extent you agree with the following statements. For validation reasons, we may have to ask similar questions.

								Not Applicable	Do Not Know
	Strongly Agree	Moderately Agree	Slightly Agree	Neutral	Slightly Disagree	Moderately Disagree	Strongly Disagree		
2. Funding Agencies' Pressure (Funding1, 2, 3, 4)								Not Applicable	Do Not Know
In my discipline,									
Data sharing is mandated by the policy of public funding agencies. (1)	1	2	3	4	5	6	7	8	9
Data sharing policy of public funding agencies is enforced. (2)	1	2	3	4	5	6	7	8	9
Public funding agencies require researchers to share data. (3)	1	2	3	4	5	6	7	8	9
Public funding agencies can penalize researchers if they do not share data. (4)	1	2	3	4	5	6	7	8	9
3. Journal Publishers' Pressure (Journal1, 2, 3, 4)								Not Applicable	Do Not Know
In my discipline,									
Data sharing is mandated by journals' policy. (1)	1	2	3	4	5	6	7	8	9
Data sharing policy of journals is enforced. (2)	1	2	3	4	5	6	7	8	9
Journals require researchers to share data. (3)	1	2	3	4	5	6	7	8	9
Journals can penalize researchers if they do not share data. (4)	1	2	3	4	5	6	7	8	9
4. Normative Pressure (Normative1, 2, 3, 4)								Not Applicable	Do Not Know
In my discipline,									
It is expected that researchers would share data. (1)	1	2	3	4	5	6	7	8	9
Researchers care a great deal about data sharing. (2)	1	2	3	4	5	6	7	8	9
Researchers share data even if not required by policies. (3)	1	2	3	4	5	6	7	8	9
Many researchers are currently participating in data sharing. (4)	1	2	3	4	5	6	7	8	9

***NOTE:** Metadata is a set of data that provides information about one or more aspects of the original research data (e.g. Ecological Metadata Language).

								Not	Do Not
	Strongly Agree	Moderately Agree	Slightly Agree	Neutral	Slightly Disagree	Moderately Disagree	Strongly Disagree		
5. Metadata* (Metadata1, 2, 3)								Not	Do Not
In my discipline,									

	Moderately Disagree								
	Strongly Disagree								
Sharing data involves too much time for me (e.g. to organize/annotate). (1)	1	2	3	4	5	6	7	8	9
I need to make a significant effort to share data. (2)	1	2	3	4	5	6	7	8	9
I would find data sharing difficult to do. (3)	1	2	3	4	5	6	7	8	9
Overall, data sharing requires a significant amount of time and effort. (4)	1	2	3	4	5	6	7	8	9
11. Data Sharing Attitude (Attitude1, 2, 3, 4)	Strongly Agree							Not Applicable	Do Not Know
	Moderately Agree								
	Slightly Agree								
	Neutral								
	Strongly Disagree								
Sharing data is valuable. (1)	1	2	3	4	5	6	7	8	9
Sharing data is desirable. (2)	1	2	3	4	5	6	7	8	9
Sharing data is pleasant. (3)	1	2	3	4	5	6	7	8	9
Sharing data is interesting. (4)	1	2	3	4	5	6	7	8	9

(Page 3)

ABOUT YOUR DATA SHARING BEHAVIOR

	Strongly Agree							Not Applicable	Do Not Know
	Moderately Agree								
	Slightly Agree								
	Neutral								
	Strongly Disagree								
12. Altruism (Altruism4, 5, 6)	Strongly Agree							Not Applicable	Do Not Know
	Moderately Agree								
	Slightly Agree								
	Neutral								
	Strongly Disagree								
I would share data to support open scientific research. (4)	1	2	3	4	5	6	7	8	9
I would share data to contribute to better scientific research. (5)	1	2	3	4	5	6	7	8	9
I would share data to help improve the quality of scientific research. (6)	1	2	3	4	5	6	7	8	9

13. In the last two years, how many publications involving actual research data have you produced per year?

(PublicationNumber)

- a) None b) 1-2 c) 3-4 d) 5-6 e) 7+

	Every time							Not Applicable	Do Not Know
	Usually								
	Frequently								
	Sometimes								
	Occasionally								
14. Data Sharing Frequencies	Occasionally							Not Applicable	Do Not Know
	Rarely								
	Never								
In the last two years, how frequently have you...	Never								
	Never								
Deposited your data into <u>disciplinary data repositories</u> for every article? (MethodDisciplinaryRepository)	1	2	3	4	5	6	7	8	9
Deposited your data into <u>institutional data repositories</u> for every article? (MethodInstitutionalRepository)	1	2	3	4	5	6	7	8	9
Uploaded your data into <u>"public" Web spaces</u> for every article? (MethodPublicWeb)	1	2	3	4	5	6	7	8	9
Provided access to your data by <u>publishing supplement materials</u> for every article? (MethodSupplement)	1	2	3	4	5	6	7	8	9
Been personally asked to share data for each article? (MethodRequested)	1	2	3	4	5	6	7	8	9
Responded to the request(s) by providing data via personal communication methods (e.g. email)? (MethodResponded)	1	2	3	4	5	6	7	8	9

15. Resources
(Resource1, 2, 3, 4, 5)

	Strongly Agree							Not	Do
	Moderately Agree								
	Slightly Agree								
	Slightly Agree							Not	Do
	Moderately Agree								
	Strongly Agree								

	Strongly Disagree			Moderately Disagree			Slightly Disagree			Neutral	
	1	2	3	4	5	6	7	8	9		
Resources that promote data sharing are available to me. (1)	1	2	3	4	5	6	7	8	9		
Specialized instruction concerning data sharing is available to me. (2)	1	2	3	4	5	6	7	8	9		
A specialized person (or group) is available to assist with my data sharing. (3)	1	2	3	4	5	6	7	8	9		
Information technologies are available to support my data sharing. (4)	1	2	3	4	5	6	7	8	9		
My organization is very supportive of my data sharing. (5)	1	2	3	4	5	6	7	8	9		

(Page 4)

ABOUT YOURSELF

16. What is your age? (Age)

- a) Under 24 b) 25-34 c) 35-44
d) 45-54 e) 55-64 f) 65+

17. What is your gender? (Gender)

- a) Male b) Female

18. What is your ethnic background? (Ethnic)

- a) Asian/Pacific Islander b) Black/African-American c) Caucasian
d) Hispanic e) Native American f) Other/Multi-Racial

19. What is your highest education so far? (Degree)

- a) Associate Degree b) Bachelor's Degree c) Master's Degree d) PhD/Doctoral Degree

20. What is your current position? (Position)

- a) Assistant Professor b) Associate Professor c) Full Professor d) Professor Emeritus
e) Professor of Practice f) Lecturer/Instructor g) Post-Doctoral Fellow h) Researcher
i) Graduate Student j) Other (Specify)

21. Please choose the option most applicable to you. (Status)

- a) Tenured b) On Tenure Track But Not Tenured c) Not on Tenure Track d) Retired

22. Which one of the following best describes your primary work sector? (Sector)

- a) Academic b) Government c) Commercial d) Non-Profit e) Other (Specify)

Once you click "SUBMIT" button below, you will be redirected to a separate page, where you can provide your email address to be entered to a drawing and receive the final report of this survey.

Please note the follow information:

- (1) Bold-face is the name of construct.
- (2) Red-text is the name of each variable.
- (3) Not Applicable and Don't Know are coded as "99" in the data.

ANEXO B – Instrumento de coleta de dados 2

Data Reuse Survey Instrument (KIM, 2017)

ABOUT YOUR DISCIPLINE

What is your primary research discipline?

DATA REUSE PERCEPTION

Please indicate to what extent you agree with the following statements. For validation reasons, we may have to ask similar questions. If you cannot answer any question(s) or they are not applicable to you, please leave them blank.

In this survey, data reuse refers to using other researchers' data for another research purpose (e.g., replicating a study, comparing to another study, conducting meta-analysis, or asking a different research question, etc.).

[Usefulness]

	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
Reusing other researchers' data:					
Improves the quality of my research	1	2	3	4	5
Enhances the effectiveness of my research	1	2	3	4	5
Reduces the time/cost/effort I spend on my research	1	2	3	4	5

[Concerns]

	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
If I reuse other researchers' data I worry that:					
I might misinterpret the data	1	2	3	4	5
I might cause infringement	1	2	3	4	5
I might not publish with that data	1	2	3	4	5

[Efforts]

	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
Reusing other researchers' data requires time and effort:					
To locate data sets	1	2	3	4	5
To access (or get permission to use) data sets	1	2	3	4	5
To process data sets for a new study	1	2	3	4	5

DISCIPLINARY CONTEXTS

[Disciplinary Climate]

	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
In my discipline:					
Researchers in my discipline cooperate well.	1	2	3	4	5
Researchers in my discipline have a strong feeling of community.	1	2	3	4	5
My discipline provides open communication among colleagues.	1	2	3	4	5

[Data Repositories]

	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
In my discipline:					

Data repositories are available for researchers to share data.	1	2	3	4	5
Researchers can easily access data repositories to reuse data.	1	2	3	4	5

YOUR ORGANIZATION & DATA REUSE INTENTION

[Organizational Resources]

In my organization (e.g., university):	Strongly Disagree				
	Disagree				
	Neutral				
	Agree				
	Strongly Agree				
Resources that promote data reuse are available to me.	1	2	3	4	5
Information technologies are available to support my data reuse.	1	2	3	4	5
A specialized person is available to assist with my data reuse.	1	2	3	4	5

[Intention to Reuse Other Researchers' Data]

	Strongly Disagree				
	Disagree				
	Neutral				
	Agree				
	Strongly Agree				
I am likely to reuse other researchers' data for my future research.	1	2	3	4	5
I intend to reuse other researchers' data for my future research.	1	2	3	4	5
I will try to reuse other researchers' data for my future research.	1	2	3	4	5

DEMOGRAPHIC INFORMATION

How many years have you worked in your current research field?

a) Less than 5 b) 6-10 c) 11-15 d) 16-20 e) 21-25 f) 26-30 g) More than 30

What is your age?

a) Under 24 b) 25-34 c) 35-44
d) 45-54 e) 55-64 f) 65+

What is your gender?

b) Male b) Female

What is your ethnic background?

a) Asian/Pacific Islander b) Black/African-American c) Caucasian
d) Hispanic e) Native American f) Other/Multi-Racial

What is your highest education so far?

a) Associate Degree b) Bachelor's Degree c) Master's Degree d) PhD/Doctoral Degree

What is your current position?

a) Assistant Professor b) Associate Professor c) Full Professor d) Professor Emeritus
e) Professor of Practice f) Lecturer/Instructor g) Post-Doctoral Fellow h) Researcher
i) Graduate Student j) Other (Specify)

Please choose the option most applicable to you.

a) Tenured b) On Tenure Track But Not Tenured c) Not on Tenure Track d) Retired

Please share any additional comments, questions, or suggestions about scientific data reuse.

Measurement Items and References

Construct	Items	Sources
Research Climate	<ul style="list-style-type: none"> • Researchers in my discipline cooperate well. • Researchers in my discipline have a strong feeling of community. • My discipline provides open communication among colleagues. 	
Availability of Data Repositories	<ul style="list-style-type: none"> • In my discipline, data repositories are available for researchers to share data. • In my discipline, researchers can easily access data repositories to reuse data. 	(Kim & Stanton, 2016; Venkatesh, Morris, Davis, & Davis, 2003)
Perceived Usefulness	<ul style="list-style-type: none"> • Reusing other researchers' data improves the quality of my research. • Reusing other researchers' data enhances the effectiveness of my research. • Reusing other researchers' data reduces the time/cost/effort I spend on my research. 	(Davis, 1989; Davis, Bagozzi, & Warshaw, 1989)
Perceived Concern	<ul style="list-style-type: none"> • If I reuse other researchers' data, I worry that I might misinterpret the data. • If I reuse other researchers' data, I worry that I might cause infringement. • If I reuse other researchers' data, I worry that I might not publish with that data. 	(Lee, 2009; Littler & Melanithiou, 2006)
Perceived Effort	<ul style="list-style-type: none"> • Reusing other researchers' data requires time and effort to locate data sets. • Reusing other researchers' data requires time and effort to access (or get permission to use) data sets. • Reusing other researchers' data requires time and effort to process data sets for a new study. 	(Davis <i>et al.</i> , 1989; Thompson, Higgins, & Howell, 1991)
Availability of Internal Resources	<ul style="list-style-type: none"> • In my organization (e.g., university), resources that promote data reuse are available to me. • Information technologies are available to support my data reuse. • A specialized person is available to assist with my data reuse. 	(Thompson <i>et al.</i> , 1991; Venkatesh <i>et al.</i> , 2003)
Data Reuse Intention	<ul style="list-style-type: none"> • I am likely to reuse other researchers' data for my future research. • I intend to reuse other researchers' data for my future research. • I will try to reuse other researchers' data for my future research. 	(Ajzen & Fishbein, 2005; Tohidinia & Mosakhani, 2010)

References

- Ajzen, I., & Fishbein, M. (2005). The influence of attitudes on behavior. In D. Albarracín, B. T. Johnson, & M. P. Zanna (Eds.), *Handbook of attitudes and attitude change* (pp. 173-221). Mahwah, NJ: Lawrence Erlbaum Associates.
- Bock, G. W., Zmud, R. W., Kim, Y. G., & Lee, J. N. (2005). Behavioral intention formation in knowledge sharing: Examining the roles of extrinsic motivators, social-psychological forces, and organizational climate. *Mis Quarterly*, 29(1), 87-111.
- Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance in information technology. *Mis Quarterly*, 13(3), 319-340.

- Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1989). User acceptance of computer technology: A comparison of two theoretical models. *Management Science*, 35(8), 982-1003.
- Kim, Y., & Stanton, J. M. (2016). Institutional and individual factors affecting scientists' data-sharing behaviors: A multilevel analysis. *Journal of the Association for Information Science and Technology*, 67(4), 776-799. doi:10.1002/asi.23424
- Lee, M.-C. (2009). Predicting and explaining the adoption of online trading: An empirical study in Taiwan. *Decision Support Systems*, 47(2), 133-142.
- Little, D., & Melanthiou, D. (2006). Consumer perceptions of risk and uncertainty and the implications for behaviour towards innovative retail services: the case of internet banking. *Journal of retailing and consumer services*, 13(6), 431-443.
- Thompson, R. L., Higgins, C. A., & Howell, J. M. (1991). Personal computing: Toward a conceptual model of utilization. *MIS Quarterly: Management Information Systems*, 15(1), 125-142.
- Tohidinia, Z., & Mosakhani, M. (2010). Knowledge sharing behaviour and its predictors. *Industrial Management & Data Systems*, 110(3-4), 611-631. doi:10.1108/02635571011039052
- Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *Mis Quarterly*, 27(3), 425-478.

ANEXO C – Instrumento de coleta de dados 3

Research data sharing and reuse (BERGHMANS *et al.*, 2017)

What is this survey about?

This survey is part of a joint project between CWTS and Elsevier to understand research data sharing and re-use practices, and the drivers (and/or barriers) acting on researchers in this regard.

This survey takes about 15-20 minutes to complete.

Interpreting the phrase ‘research data’ in this survey

Research data is defined as ‘recorded factual material generated (and commonly retained) and accepted in the research community as necessary to derive and validate research findings’.

Q1a. Categories of research data used

Please select the categories of research data that you used in your last research project:

- Observational or empirical: e.g. sensor data, survey data, interview transcripts, sample data, neuroimages, ethnographic data, diaries.
- Experimental: e.g. gene sequences, chromatograms, toroid magnetic field data.
- Simulation: e.g. climate models, economic models.
- Derived or compiled: e.g. text and data mining, compiled database, 3D models.
- Reference or canonical: e.g. gene sequence databanks, chemical structures, spatial data portals.

Q1b. Research data format(s) used

Please select the research data formats that you used or created as part of your last research project:

- Text: e.g. flat text files.
- Numerical: e.g. SPSS, Stata, Excel.
- Databases: e.g. MS Access, SQL Server, MySQL, Oracle.
- Multimedia: e.g. jpeg, tiff, mpeg, QuickTime, mp3, DICOM.
- Models: e.g. 3D, statistical.
- Software: e.g. Java, C.
- Discipline specific: e.g. FITS in astronomy, CIF in chemistry.
- Instrument specific: e.g. proprietary microscope formats such as Olympus Confocal Microscope Data Format, Carl Zeiss Digital Microscopic Image Format (ZVI).

Q1c. Research data format(s) to be shared

Which of the following research data formats you have indicated that you used or created as part of your last research project are important to be made available to others for the purpose of validation?

[Populate with choices from previous data formats question]

- Text: e.g. flat text files.
- Numerical: e.g. SPSS, Stata, Excel.
- Databases: e.g. MS Access, SQL Server, MySQL, Oracle.
- Multimedia: e.g. jpeg, tiff, mpeg, QuickTime, mp3, DICOM.
- Models: e.g. 3D, statistical.
- Software: e.g. Java, C.
- Discipline specific: e.g. FITS in astronomy, CIF in chemistry.
- Instrument specific: e.g. proprietary microscope formats such as Olympus Confocal Microscope Data Format, Carl Zeiss Digital Microscopic Image Format (ZVI).

Q1c1. File size

Approximately, what is the volume of the data you created as part of your last research project?

- Less than 1 gigabyte
- 1-9 gigabytes
- 10-49 gigabytes
- 50-99 gigabytes
- 100-1000 gigabytes (up to 1 terabyte)
- 1-10 terabytes
- More than 10 terabytes
- Don't know

Q1c2. Number of data files

Approximately how many files of data did you produce as part of your last research project?

- Fewer than 10

	Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree	Don't know
I encounter obstacles sharing my research data with others	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sharing research data is associated with credit or reward in my field	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sharing research data is important for doing research in my field	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I have received sufficient training in research data sharing	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I provide my research data to publishers so that it can be made accessible with my research article	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Research data management specialists need to play a role in research data sharing	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q2b. Benefits of access to others' research data

[If response to Q2a. 'Having access to others' research data benefits/would benefit my own research' = Agree or Strongly Agree]

How has (or would) accessing others' research data benefitted you?

Where have (or would) you accessed others' research data? Please list specific websites, for example.

Q2c. Benefits of publishing research data alongside article

Which, if any of the following, do you think are benefits to you as an author for sharing your research data alongside your research article?

[If response to Q2a. 'I provide my research data to publishers so that it can be made accessible with my research article' = Agree or Strongly Agree]

- Article more likely to be accepted for publication
- Article more likely to be cited
- Compliance with funding body mandates
- Compliance with journal or publisher requirements
- Reproducibility of research
- Data reuse
- Research aggregation (e.g. for meta-analysis)
- More possibilities for collaboration
- Encourage other researchers to make their data publicly available
- Other, please specify _____
- Don't know [exclusive]

Q2d. Drawbacks of not publishing research data alongside article

Using the space below, please indicate what, if any, would be the drawback for you as the author if you were to share your research data with your research article?

[If response to Q2a. 'I provide my research data to publishers so that it can be made accessible with my research article' = Disagree or Strongly Disagree]

Q2e. Research data management and archiving

Do you take steps to manage your research data and/or archive it for potential re-use by yourself and/or others?

- Yes- always
- Sometimes
- Never

Q2f. Approach to research data management

[If response to Q2e. 'Do you take steps to manage your research data and archive it for potential re-use by yourself and/or others?' = Yes or sometimes]

How would you describe your approach to research data management?

- Very structured
- Structured
- Unstructured
- Very Unstructured
- Unsure
- I don't have any data to manage

Q2g. Research data sharing locations

[If response to Q2e. 'Do you take steps to manage your research data and archive it for potential re-use by yourself and/or others?' = Yes or sometimes]

Which of the following locations do you use to archive your research data?

- Repository provided by my funder
- Repository provided by a publisher
- Repository provided by my institute
- Repository provided by my department
- I don't archive my research data
- Other, please specify _____

Q2h. Research data sharing actions

[If response to Q2g. 'Which of the following locations do you use to archive your research data?' = Any answer except 'I don't archive my research data']

Who usually does the archiving of your research data at these locations?

- Someone else on my behalf (e.g. repository manager)
- Myself
- Other, please specify _____

Q2i. Drivers of research data management

[If response to Q2e. 'Do you take steps to manage your research data and archive it for potential re-use by yourself and/or others?' = Yes]

Which of the following impacts your decision to archive research data?

	No impact	Some impact	Great impact	Mandates archiving
Funder	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Publisher	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Institute (e.g. Library, Research Office, IT Department, etc.)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
My department (e.g. Department Head and/or Supervisor)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Colleagues/collaborators	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
My wider research field (e.g. documented code of conduct or informal practice)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Myself (i.e. own initiative)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Other, please specify _____	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Q2j. Role of research data management plans

[If response to Q2i. 'Which of the following impacts your decision to archive research data?' = Mandates archiving for at least one category, randomly assign the following questions]

In your previous response you are mandated to archive your research data by the following category: [name of category]. Do they have a research data management plan that you follow?

- Yes
- No
- Don't know

Q2k. Implementation of research data management plans

[If response to Q2j. 'In your previous response you are mandated to archive your research data by the following category: [name of category]. Do they have a research data management plan that you follow?' = Yes]

Who is responsible for the execution this research data management plan?

- Funder's research data management organisation
- Institutional Library (e.g. Librarian)
- Institutional Research Office (e.g. Research Administrator)
- Institutional IT Department
- My department (e.g. Department Head and/or Supervisor)
- Other, please specify _____
- Don't know [EXCLUSIVE]

Q2l. Compliance with research data management plans

[If response to Q2j. 'In your previous response you are mandated to archive your research data by the following category: [name of category]. Do they have a research data management plan that you follow?' = Yes]

Who is responsible for monitoring compliance this research data management plan?

- Funder's research data management organisation
- Institutional Library (e.g. Librarian)
- Institutional Research Office (e.g. Research Administrator)
- Institutional IT Department
- My department (e.g. Department Head and/or Supervisor)
- Other, please specify _____
- Don't know [EXCLUSIVE]

Q2m. Funding in place

[If response to Q2e. 'Do you take steps to manage your research data and/or archive it for potential re-use by yourself and/or others?' = Yes]

Does your institute allocate funds to cover the costs of managing and/or archiving research data?

- Yes
- No
- Don't know

Q2n. Funding growth

[If response to Q2m. 'Does your institute allocate funds to cover the costs of managing research data?' = Yes]

Thinking about the funding your institute spent on managing and archiving research data, do you expect the funding in 2016 to increase, remain the same or decrease when compared to 2015? [Single-code]

- Increase
- Remain the same
- Decrease
- Don't know

Q2o. Funding change

(IF Q2n= 2 Punch "0" IF Q2n = 1 OR 3.)

Approximately, by what percentage do you expect it to [INSERT increase / decrease FROM Q2m)?
_____ %

Q3a. Research data I have shared directly

Have you shared directly (i.e. person-to-person sharing) with others any or all of the research data that you used or created as part of your last research project?

- Yes
- No
- Don't know

Q3b. Groups research data shared with directly

[If response to Q3a. 'Have you shared directly with others any or all of the research data that you used or created as part of your last research project?' = Yes]

Irrespective of public sharing (e.g. archiving your research data in a repository accessible to others), have you done any of the following with any or all of the research data that you used or created as part of your last research project?

- Shared directly with researchers working on the same research project in a research collaboration
- Shared directly with researchers NOT working on the same research project who I know personally
- Shared directly with researchers NOT working on the same research project who I DON'T know personally
- Shared directly with research project partners (e.g. funders)
- Other, please specify _____

Q3c. Research data shared with me by others

Was any or all of the research data that you used or created as part of your last research project shared with you by others, either directly (i.e. person-to-person sharing) or indirectly (i.e. downloaded from a repository)?

- Yes
 No
 Don't Know

Q3d – Whether will use Creative Commons

Thinking your next set of research data, if you were to make it available under which license(s) would you make it available?

[NOTES explanation in brackets to appear as hovers]

- CC 0 (no rights reserved by the creator)
 CC BY Attribution (allows others, even commercially, to distribute and modify your work as long as they credit you for the original creation)
 CC BY-NC NonCommercial (allows non-commercial use but others do not have to license derivative work on the same terms)
 CC BY-NC-ND Attribution-NonCommercial-NoDerivs (others can download your work and share, as long as they credit you, but cannot modify or use your work commercially)
 Other, please specify _____
 Don't know

Q3e. Trust in others' research data

[If response to Q3c. 'Was any or all of the research data that you used or created as part of your last research project shared with you by others, either directly (i.e. person-to-person sharing) or indirectly (i.e. downloaded from a repository)?' = Yes]

How important were the following factors when deciding to use others' research data that the data...

[Randomise statements. Single code each statement]

	Very important	Important	Somewhat important	Unimportant	Not at all important	Don't know
Are from someone I know personally	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Are from someone I DON'T know personally but who has a good reputation	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Are from someone at an institute with a good reputation	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Are from a repository with a good reputation	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Comply with guidelines around collection and/or formatting	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Have good documentation and/or metadata	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Are certified or accredited by a third party	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Are commonly used by others	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Have been cited in a publication in the formal literature (journals, books, etc.)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q3f. Citing others' research data

Publications in the formal literature (journals, books, etc.) typically make reference to (cite) other published literature, usually in a separate cited references section. There is an increasing need to also cite research data that is not itself included in the formal literature but has been shared directly (i.e. person-to-person sharing) or indirectly (i.e. downloaded from a repository). [SINGLE CODE]

Please select from the following statements about standards for citing research data.

- There are no agreed standards for citing research data in my field
 There are agreed standards for citing research data in my field, but most researchers follow them

- There are agreed standards for citing research data in my field, but most researchers do not follow them
- There is no need for to have standards for the citation of research data

Q3g. Effort required prior to sharing data

Prior to sharing, research data typically needs to be formatted and often has documentation and/or metadata added to make it re-useable by others. How would describe the effort typically required to make your research data re-useable by others?

- A lot of effort
- Some effort
- Little effort
- No effort
- I don't share my research data

Q3g. Why Effort

[If response to Q3c. 'Prior to sharing, research data typically needs to be formatted and often has documentation and/or metadata added to make it re-useable by others. How would describe the effort typically required to make your research data re-useable by others?' = A lot of effort]

You said that it takes a lot of effort to make your research data re-useable, what is the main reason for this?

DEMOGRAPHIC SECTION

You are nearly at the end of the survey. Some questions to help us classify your answers.

P1 - Country

In which country are you based?

Afghanistan

to

- Zimbabwe

- P2 - Subject

In which subject discipline do you specialize?

Agriculture

- Arts and Humanities

- Astronomy

- Biochemistry, Genetics, and Molecular Biology

- Biological Sciences

- Chemical Engineering

- Chemistry

- Computer Sciences / IT

- Earth and Planetary Sciences

- Economics

- Electrical / Electronic Engineering

- Engineering and Technology

- Environmental Sciences

- Immunology and Microbiology

- Materials Science

- Mathematics

- Medicine and Allied Health

- Neuroscience

- Nursing

- Pharmacology, Toxicology and Pharmaceutics

- Physics

- Psychology

- Social Science

- Other (please specify)_____

P2a - Medical/allied health specialty

Which one of the following fields of medicine / allied health are you mainly involved in?

- Allergy, Asthma & Immunology

- Anesthesiology & Pain Management

- Cardiology

- Critical Care

- Dentistry

- Dermatology

- Emergency Medicine

- Endocrinology & Metabolism

- General / Family Medicine

- Gastroenterology

- Geriatric Medicine

- Hematology

- Imaging technology

- Infectious Disease

- Internal Medicine
- Laser Vision Surgery
- Medical Genetics
- Nephrology
- Neurology
- Neurosurgery
- Obstetrics & Gynecology
- Occupational & Environmental Medicine
- Oncology
- Ophthalmology
- Orthopedic Surgery
- Otorhinolaryngology
- Osteopathy
- Pathology
- Pediatrics
- Physical Medicine & Rehabilitation
- Physical Therapy
- Plastic Surgery
- Preventive Medicine
- Proctology
- Psychiatry
- Pulmonary Medicine
- Radiology (including Radiology Technicians)
- Rheumatology
- Sports Medicine
- Surgery, General
- Thoracic Surgery
- Urology
- Veterinary Medicine
- Other Specialties

P3 - Number articles published in career

Approximately how many articles have you published so far in your career?

- 1
- 2-5
- 6-15
- 16-25
- 26-50
- 51-75
- 76-100
- 101-200
- Over 200
- None

P5 - Organization

What type of organization do you work for?

- University or college
- Research Institute
- Medical School/ hospital
- Corporate
- Consultant
- Other (Please specify)_____

P6 - Position

What is your position within your organization?

- Head of Department/ Senior Management

- Senior Researcher/ Middle Management
- Researcher/ Staff Member
- Other (please specify)_____

P7 - Main role

What is your MAIN role within your organization?

- Research and/or development
- Teaching
- R&D and teaching equally
- Management/Administration
- Practitioner (clinical)
- Practitioner (engineering/technology)
- Advisory/Consultancy
- Other (please specify)_____

P8 - Medical profession only

Which of the below best describes your profession?

- Resident
- Fellow
- Physician
- RN
- Pharmacist
- Physician Assistants

P9 - Gender

Which gender are you?

- Male
- Female
- Other
- Prefer not to say

P10 - Age

Please indicate your age group:

- Under 26
- 26-35
- 36-45
- 46-55
- 56-65
- Over 65
- Prefer not to say

STOP

Complete – Complete

The survey is now complete and your results have been submitted. Thank you for taking part in this important study. You may now close this window.

ÍNDICE ONOMÁSTICO

ACRL INFORMATION LITERACY BEST PRACTICES COMMITTEE, 2012	95
ADAMICK; REZNIK-ZELLEN; SHERIDAN, 2012	101, 298
ALBAGLI, 2004	49, 50
ALBAGLI, 2015	49
ALVES, 2016	43, 162
AMERICAN LIBRARY ASSOCIATION, 1989	24
AMORIM <i>et al.</i> , 2015	20
AMORIM, 2013	43, 162
ANDRADE, 2018	47
APPEL, 2014	43, 162
ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS, 2002	167
ASSOCIATION OF COLLEGE AND RESEARCH LIBRARIES, 2000	94, 95, 96, 133
ASSOCIATION OF COLLEGE AND RESEARCH LIBRARIES, 2012a	95
ASSOCIATION OF COLLEGE AND RESEARCH LIBRARIES, 2012b	106
ASSOCIATION OF COLLEGE AND RESEARCH LIBRARIES, 2013	92, 95, 97, 255, 260
ASSOCIATION OF COLLEGE AND RESEARCH LIBRARIES, 2014	23, 298
ASSOCIATION OF COLLEGE AND RESEARCH LIBRARIES, 2015	23, 95
ASSOCIATION OF RESEARCH LIBRARIES, [s.d.]	104
ASSOCIATION OF RESEARCH LIBRARIES, 2006	100
ASSOCIATION OF RESEARCH LIBRARIES, 2010	100
ASSOCIATION OF RESEARCH LIBRARIES, 2012	79
AVENTURIER; ALENCAR, 2016	29
AWRE <i>et al.</i> , 2015	93
BALL, 2012	99
BARDIN, 1977	140
BARDIN, 1995	52, 133, 141, 186, 193, 194
BARDIN, 2011	140
BAWDEN; ROBINSON, 2009	93
BERGHMANS <i>et al.</i> , 2017. 132, 134, 138, 139, 141, 144, 146, 187, 200, 201, 202, 203, 206, 211, 222, 225, 227, 232, 290, 301, 415	
BERLIN, 2003	22
BERMAN, 2013	48
BILL & MELINDA GATES FOUNDATION, 2015	29
BINGERT; BUDDENBOHM, 2016	102, 298
BIOTECHNOLOGY AND BIOLOGICAL SCIENCE RESEARCH COUNCIL, 2008	50
DOCUMENTATION AND LIBRARY DIVISION OF THE FACULTY OF SCIENCES AND TECHNOLOGY, 2018	86
BORGMAN <i>et al.</i> , 2015	20
BORGMAN, 2000	22, 23, 24, 38, 54, 132, 133, 136
BORGMAN, 2007	21, 22, 24, 25, 38, 54, 58, 61, 62, 69, 70, 75, 76, 124, 132, 133, 136, 260
BORGMAN, 2012	21, 68
BORGMAN; FURNER, 2012	61
BOSCH; HENDERSON, 2012	79
BRASIL, 2007	81
BRASIL, 2009	82
BRASIL, 2011	30, 81
BRASIL, 2012a	81
BRASIL, 2012b	81

BRASIL, 2013a.....	82
BRASIL, 2013b.....	82
BRASIL, 2013c.....	82
BRASIL, 2014.....	82
BRASIL, 2014a.....	82
BRASIL, 2015.....	153
BRASIL, 2017a.....	31
BRASIL, 2017b.....	29
BRAZILIAN POLITICAL SCIENCE REVIEW, [s.d.]	47
BRINEY; GOBEN; ZILINSKI, 2015	29, 83, 85
BROWN, 2013.....	58, 70, 88, 133, 298
BRUCE <i>et al.</i> , 2013.....	114, 115
BRUCE, HUGHES, SOMERVILLE, 2012	115
BRUCE; HUGHES, 2010	115
BRYANT; LAVOIE; MALPAS, 2017	33
BUENO DE LA FUENTE, 2016	31
CALLON, 1986.....	70, 71, 134
CALLON, 1993.....	134
CALLON; LAW, 1989.....	61, 306
CALZADA PRADO; MARZAL, 2013.....	40, 96, 98, 103, 108, 110, 113, 132, 133, 144, 187, 189, 261, 300, 354, 365
CANADA, 2004.....	149, 150, 151, 152
CARLSON <i>et al.</i> , 2011 .	24, 25, 33, 48, 53, 94, 95, 96, 97, 101, 105, 108, 109, 113, 118, 132, 133, 144, 187, 189, 223, 254, 255, 261, 298, 300, 354
CARLSON; JOHNSTON, 2015.....	24, 94, 102, 108, 109, 112, 115, 118, 132, 133, 144, 187, 189, 223, 254, 255, 261, 298, 300, 354
CARTER, 2013	58, 69, 70, 133
CARVALHO, 2016.....	31, 46, 47
CARVALHO, 2018.....	43, 162
CARVALHO; RODRIGUES; PRÍNCIPE, 2014	110
CENDI, 2016.....	29, 83
CESAR JUNIOR, 2011	60
CHRISTENSEN-DALSGAARD <i>et al.</i> , 2012	106, 298
COELHO, 2017.....	43, 162
CONSORTIUM OF EUROPEAN SOCIAL SCIENCE DATA ARCHIVES, 2018.....	91, 92, 120, 134, 187, 207, 354
COPE; KALANTZIZ	31
CORTI <i>et al.</i> , 2014.....	121, 134, 187, 208
COSTA, 2017.....	43, 162
COUNCIL ON LIBRARY AND INFORMATION RESOURCES, 2008.....	100
COUNCIL ON LIBRARY AND INFORMATION RESOURCES, 2013.....	98
CREASER, 2011	21, 22, 23, 34, 48, 79, 87, 298
CULLEN; CHAWNER, 2011	76, 87, 298
CURATING RESEARCH ASSETS AND DATA USING LIFECYCLE EDUCATION (CRADLE); UNIVERSITY OF NORTH CAROLINA; UNIVERSITY OF EDINBURGH, 2017	134, 187, 208
CURTY <i>et al.</i> , 2017	24, 39, 40, 54, 134, 136, 213
CURTY, 2016a.....	21, 24, 37, 54, 132, 133, 136, 213
CURTY, 2016b	21, 24, 54, 132, 133, 136
CUTCLIFFE, 2003.....	48, 62, 260
DATAONE, 2012.....	117, 120, 134, 187, 207
DAVIS <i>et al.</i> , 2008.....	78, 132, 133
DECHMAN; SYMS, 2014.....	101, 298

DELANEY; BATES, 2015	40, 94, 96, 100, 298
DIGITAL CURATION CENTER, 2007.....	90
DIGITAL CURATION CENTER, 2014.....	121, 134, 207
DIGITAL SCIENCE; FIGSHARE, 2016.....	32
DULCINEA, <i>online</i>	86
DURAND, 1998.....	108
DURAND, 2000.....	108
ELLWEIN <i>et al.</i> , 2014	101, 115, 298
ELSEVIER, 2013	102
ESTEVÃO, 2019.....	299
EUROPEAN COMMISSION, 2006	50, 79
EUROPEAN COMMISSION, 2014	37, 78
EUROPEAN COMMISSION, 2016a.....	20, 29, 37, 215
EUROPEAN COMMISSION, 2016b	37
EUROPEAN COMMISSION, 2016c.....	20, 38, 58
EUROPEAN COMMISSION, 2017	59
EYNDEN; SCHRECKENBERG; CORTI, 2015	91
FACILITATE OPEN SCIENCE TRAINING FOR EUROPEAN RESEARCH, 2015 .	58, 94, 108, 110, 113, 121, 132, 133, 134, 144, 187, 189, 208, 223, 255, 261, 300, 354, 368
FANIEL <i>et al.</i> , 2013.....	31
FANIEL; MAUTHNER, 2012	31
FECHER; FRIESIKE, 2014	49, 59, 60, 77, 125, 132, 133
FEDERER <i>et al.</i> , 2016	101, 298
FEENBERG, 2010	49, 62
FIGUEIREDO, 2018.....	43, 162
FLORES, 2006.....	50
FONG; WANG, 2015.....	101, 102, 298
FRANK; PHARO, 2016.....	48, 101, 298
FRANKEL <i>et al.</i> , 2000.....	124
FUNDAÇÃO DE AMPARO À PESQUISA DO ESTADO DE SÃO PAULO, 2018	31
GARVEY, 1979	21
GIARLO, 2013.....	98, 254
GIL, 2010	137, 141
GITBOOK BOT <i>et al.</i> , 2018	59
GRAY, 2005.....	24
GRIGG, 2015	23, 28, 71
GUEDON, 1994	75
GUERTIN; BERNHARD, 2005.....	149, 150, 151, 152
HEIDORN, 2011	107
HELBIG, 2016	102, 298
HEY; TANSLEY; TOLLE, 2009.....	61, 66, 89, 107, 128
HEY; TREFETHEN, 2003.....	61, 92
HUGHES, 1993.....	70
INDUSTRY CANADA, 2014.....	50
INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS, 2014.....	61
INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA, 2005	80
INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA, 2009.....	80
INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA, 2016.....	80
INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA, 2018.....	86
INTER-UNIVERSITY CONSORTIUM FOR POLITICAL AND SOCIAL RESEARCH, 2012	99, 100, 132, 133
INTER-UNIVERSITY CONSORTIUM FOR POLITICAL AND SOCIAL RESEARCH, 2017	27
INTER-UNIVERSITY CONSORTIUM FOR POLITICAL AND SOCIAL RESEARCH, 2019	185

JAHNKE; ASHER; KERALIS, 2012	33, 53, 107, 298
JEFFRYES; JOHNSTON, 2013.....	24, 33, 101, 108, 112, 132, 133, 144, 187, 189, 261, 298, 300, 354
JISC SERVICES LIMITED, 2011	90
JOHNSON, 2012.....	24
JOHNSON; BRESNAHAN, 2015	101, 298
JOHNSTON; JEFFRYES, 2012.....	120, 134, 187, 207
JORDAN, 1973	21
JORGE, 2018	43, 162
JUBB, 2013	61
JUBB; SHORLEY, 2013.....	22
KAFEL; CREAMER; MARTIN, 2014.....	106, 117, 118, 127
KARDELIENE; KARDELIS; BAKUTYTE, 2014.....	88, 298
KARON <i>et al.</i> , 2013.....	106, 107
KILEY, 2013.....	58, 69, 70, 77, 78, 79, 133, 215
KIM, 2013	21, 22, 212
KIM, 2016.....	27, 132, 134, 138, 139, 141, 144, 146, 187, 200, 202, 203, 206, 211, 212, 216, 218, 219, 222, 232, 290, 301, 407
KIM, 2017.....	27, 132, 134, 138, 139, 141, 144, 146, 187, 200, 202, 203, 206, 211, 219, 221, 222, 232, 290, 301, 393, 411
KIMBROUGH, 2016	21
KLING, 2004	75
KOLTAY, 2015a.....	48, 54, 92, 94, 132, 133
KOLTAY, 2015b.....	103, 298
KOLTAY, 2016.....	24, 25, 48, 54, 132, 133
KOLTAY, 2017a.....	24, 33, 48, 54, 107, 132, 133
KOLTAY, 2017b.....	48, 88, 102, 103, 104, 132, 133, 298
KOLTAY, 2017c.....	93
KOLTAY; HORSTMANN; WITT, 2016	20, 94, 103, 298
KOLTAY; SPIRANEC; KARVALICS, 2016	92, 98
KURAMOTO, 2006.....	80
KURAMOTO, 2009.....	80, 82
KURAMOTO, 2013.....	82
KURAMOTO, 2014.....	82, 83
LAKATOS; MARKONI, 2012	141
LAKATOS; MARKONI, 2017	51, 141
LANCASTER, 1977	21
LATOUR, 1986.....	64, 134
LATOUR, 1992.....	27, 66, 134
LATOUR, 1993.....	134
LATOUR, 1995.....	134
LATOUR, 1998.....	123
LATOUR, 2000.....	35, 38, 39, 63, 64, 67, 76, 123, 124, 125, 134, 260, 263, 300
LATOUR, 2001.....	35, 54, 63, 64, 66, 67, 69, 123, 134, 260
LATOUR, 2006.....	63, 65, 134
LATOUR; CALLON, 1981.....	38, 123, 124, 134
LEIDEN UNIVERSITY; ELSEVIER, 2017.....	27, 31, 32, 139, 144, 187, 222
LEONELLI, 2013.....	31
LIEVROUW, 1990.....	21, 24
LINSINGEN; BAZZO; PEREIRA, 2003.....	49, 61, 62, 306
LÓSCIO; BURLE; CALEGARI, 2017	134, 135, 180, 291, 292
LUNA, 2000.....	142
LYMAN, 1999	22

MACHADO, 2015	43, 74, 162
MACY; COATES, 2016	48, 101, 298
MAIA, 2018	43, 162
MANDINACH; GUMMER, 2013	48, 97, 101, 298
MANDINACH; GUMMER, 2015	48
MARCONI; LAKATOS, 2007	133, 145
MARCONI; LAKATOS, 2012	133, 137, 138, 139
MARQUES, 2014	43, 162
MASSACHUSETTS INSTITUTE OF TECHNOLOGY, 2014.....	101, 298
MATTAR, 2014	116, 126, 296, 307
MAYBEE <i>et al.</i> , 2015	24, 33
MAYBEE; ZILINSKI, 2015	113, 114, 115
MAYERNIK <i>et al.</i> , 2014.....	105, 298
MCGRATH, 2013	58, 69, 70, 133
MEADOWS, 1974	21, 53
MEADOWS, 1998	21, 53, 75, 76
MEDEIROS, 2015.....	43, 44, 162, 163
MEDEIROS; CAREGNATO, 2012.....	60, 61
MEDICAL RESEARCH COUNCIL, 2006	50
MEES, 1918	76, 77
MONTEIRO, 2017.....	21, 43, 163
MOONEY <i>et al.</i> , 2014	101, 298
NATIONAL ENDOWMENT FOR THE HUMANITIES, 2018	90
NATIONAL INSTITUTES OF HEALTH, 2003	28, 50
NATIONAL SCIENCE BOARD, 2005	20, 22, 50, 89, 90, 91
NATIONAL SCIENCE FOUNDATION, 2007.....	20, 23, 29, 50, 215
NATURE PUBLISHING GROUP, 2014.....	85
NECDMC, 2012.....	134, 187, 208
NEW MEDIA CONSORTIUM, 2013.....	40, 96
NOVELLI, 2012.....	43, 162
OAKLIST, <i>online</i>	86
OECD, 2004	50
OFFICE OF SCIENCE AND TECHNOLOGY POLICY, 2013.....	29, 50
OLIVEIRA, 2016	43, 162
OPEN AIRE, 2015	78
OPEN AIRE, 2017	285
OPEN RESEARCH FUNDERS GROUP, 2016	29, 37
OPEN SCIENCE AND RESEARCH INITIATIVE, 2014	79
ORGANIZATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT, 2007.....	23
ORGANIZATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT, 2015.....	20, 37, 58
ORTEGA; GASSET, 1967.....	87
PAGANI; KOVALESKI; RESENDE, 2017	149
PAISLEY, 1984.....	21, 24
PEPE <i>et al.</i> , 2009.....	76, 77
PEREIRA; SCHMITT; DIAS, 2007	28, 116, 117, 126, 127, 128, 306, 307
PIWOWAR; VISION, 2013	126
PLOS ONE, [s.d.].....	85
PONTIKA <i>et al.</i> , 2015	25
POOLE, 2016.....	21
PRADO, 2017	43, 162
PRAHALAD; HAMEL, 1990.....	109
PROSSER, 2013.....	57, 58, 70, 133

PUBLIC LIBRARY OF SCIENCE, [s.d.]	47, 85
QIN; D'IGNAZIO, 2010	24, 94, 298
RAFFESTIN, 1993.....	49, 50
REDE NACIONAL DE ENSINO E PESQUISA, 2017a.....	30
REDE NACIONAL DE ENSINO E PESQUISA, 2017b.....	46
RESEARCH DATA ALLIANCE, 2017	30, 37
RESEARCH DATA CANADA, 2016	50
RESEARCH DATA NETHERLANDS, <i>online</i>	134, 207
RILEY; TREVOR, 2016	98, 117, 118, 254
RITTEL; WEBBER, 1973.....	93
ROARMAP, 2018.....	83, 84
ROCHA, 2017.....	30, 46
RODRIGUES, 2008	66
ROLANDO; BORGMAN, 2013	89
ROSENDAAL; GEURTS, 1997	64, 65, 71, 75, 123, 125, 126, 133
ROYAL NETHERLANDS ACADEMY OF ARTS AND SCIENCES, 2018.....	39
ROYAL SOCIETY, 2012	90, 91, 254
SALES; SAYÃO, 2015a	21
SAMPIERI; COLLADO; LUCIO, 2013	142, 144
SANTOS, 2018	43, 162
SAYÃO, 2012	20
SAYÃO; MARCONDES, 2009	80
SAYÃO; SALES, 2012.....	22, 53, 64
SAYÃO; SALES, 2014.....	24, 30, 39, 47, 113, 213
SAYÃO; SALES, 2015a.....	39
SAYÃO; SALES, 2015b.....	20, 21, 38
SAYÃO; SALES, 2016.....	35, 53, 60, 100
SAYES, 2014	74, 75
SCHNEIDER, 2013.....	24, 92, 94, 95, 98, 133, 260
SCHONFELD, 2013	58, 70, 133
SCIENCE-METRIX, 2018.....	79, 80
SETENARESКИ; SUNYE; SHIMA, 2013	81
SHERPA RoMEO, <i>online</i>	86, 223
SHORISH, 2015.....	96
SILVA, 2017	43, 162
SISMONDO, 2010.....	70
SOCIAL SCIENCES AND HUMANITIES RESEARCH COUNCIL OF CANADA, 1990	50
SOUSA, 2014.....	43, 162
SOUZA, 2016.....	43, 162
SURKIS <i>et al.</i> , 2017.....	101, 298
TASK GROUP ON DATA CITATION STANDARDS AND PRACTICES, 2013.....	126, 223, 255
TENOPIR <i>et al.</i> , 2013	88, 100, 104, 106, 298
TENOPIR <i>et al.</i> , 2014.....	102, 103, 107, 254, 298
TENOPIR <i>et al.</i> , 2015a	23, 32, 33, 40, 87, 88, 298
TENOPIR <i>et al.</i> , 2015b	47, 100, 104, 105, 106, 107
TENOPIR <i>et al.</i> , 2016.....	47, 53, 88, 100, 104, 105, 298
TENOPIR <i>et al.</i> , 2017	47, 87, 93, 100, 104, 105, 106, 298
TENOPIR, 2011	24, 54, 132, 133, 136, 213
TENOPIR, 2015	24, 54, 132, 133, 136
TENOPIR, 2016.....	105
TENOPIR; BIRCH; ALLARD, 2012.....	21, 23, 47, 87, 88, 99, 101, 104, 105, 106, 298
TENOPIR; HUGHES; ALLARD, 2015.....	88, 298

TENOPIR; KING, 2009	69
THE ENGINEERING AND PHYSICAL SCIENCES RESEARCH COUNCIL, 2018	89, 132, 133
THE WELCOME TRUST, 2007.....	50
TRI-AGENCY, 2014.....	50
TYGEL, 2016.....	43, 162
UK RESEARCH COUNCIL, 2005.....	50
UNITED NATIONS EDUCATIONAL, SCIENTIFIC AND CULTURAL ORGANISATION, 2015	139
UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ, [s.d.]	48
UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ, 2018	48
UNIVERSITY OF BATH, 2011	90
UNIVERSITY OF EDINBURGH, 2015.....	102
UNIVERSITY OF EDINBURGH, 2016.....	90
UNIVERSITY OF EDINBURGH, 2017.....	118, 121, 134, 187, 208
UNIVERSITY OF MINNESOTA, 2014.....	101, 298
UNIVERSITY OF OXFORD, 2018.....	134, 187, 207
VACCAREZZA, 2011	48
VARVEL; BAMMERLIN; PALMER, 2012	106
VEIGA, 2017	30
VERBAAN; COX, 2014	102, 298
VITAE, 2011	108, 111, 113, 132, 133, 144, 187, 189, 261, 300, 354
WALLIS; ROLANDO; BORGMAN, 2013.....	90
WANNER, 2015	95, 96, 97, 298
WEB OF SCIENCE, 2017	153
WHITLEY, 1984.....	24
WHITMIRE, 2015.....	98, 99, 298
WHYTE <i>et al.</i> , 2014.....	21
WITT <i>et al.</i> , 2017.....	101
WITT, 2012.....	94, 101, 298
WOOD, 2013	77, 254
ZIMMERMAN, 2011.....	31
ZOTERO, 2018	165