



UnB

Universidade de Brasília – UnB

Faculdade de Ciência da Informação – FCI

Programa de Pós-Graduação em Ciência da Informação

**FATORES QUE INFLUENCIAM A COMUNICAÇÃO DE DADOS DE PESQUISA SOBRE
O VÍRUS DA ZIKA NA PERSPECTIVA DE PESQUISADORES**

TESE DE DOUTORADO

Michelli Costa

Orientador: Fernando César Lima Leite

Brasília,
Fevereiro de 2017

MICHELLI COSTA

**FATORES QUE INFLUENCIAM A COMUNICAÇÃO DE DADOS DE PESQUISA SOBRE
O VÍRUS DA ZIKA NA PERSPECTIVA DE PESQUISADORES**

Tese apresentada ao Programa de Pós-Graduação em Ciência da Informação (PPGCIInf) da Universidade de Brasília (UnB) como requisito parcial para obtenção de título de Doutora em Ciência da Informação, sob a orientação do Prof. Dr. Fernando César Lima Leite.

Brasília

Fevereiro de 2017

FOLHA DE APROVAÇÃO

Título: "Fatores que influenciam a comunicação de dados de pesquisa sobre o vírus da zika, na perspectiva de pesquisadores".

Autor (a): Michelli Pereira da Costa

Área de Concentração: Gestão da Informação

Linha de pesquisa: Comunicação e Mediação da Informação

Tese submetida à Comissão Examinadora designada pelo Colegiado do Programa de Pós-graduação em Ciência da Informação da Faculdade em Ciência da Informação da Universidade de Brasília como requisito parcial para obtenção do título de **Doutor** em Ciência da Informação.

Tese aprovada em: 13 de fevereiro de 2017.



Prof. Dr. Fernando César Lima Leite (UnB/PPGCINF)



Prof. Dr.ª Sely Maria de Souza Costa (UnB/PPGCINF)



Prof. Dr.ª Rosemeire Barbosa Tavares (JBICT)



Prof. Dr.ª Patrícia Rocha Bello Bertin (EMBRAPA)



Prof. Dr.ª Ana Valéria Machado Mendonça (PPGSC/UnB)

Prof. Dr.ª Marília Augusta de Freitas (BCE/UnB)

C837f

Costa, Michelli Pereira da

Fatores que influenciam a comunicação de dados de pesquisa sobre o vírus da zika, na perspectiva de pesquisadores. / Michelli Pereira da Costa; orientador Fernando César Lima Leita. – Brasília: Universidade de Brasília, 2017.

Tese (Doutorado – Doutorado em Ciência da Informação) – Universidade de Brasília, 2017.

1. comunicação científica. 2. Dados de pesquisa. 3. Ciência aberta. 4. Vírus da zika. I. Leite, Fernando César Lima, orient. II. Título.

Agradecimentos

Agradeço de forma muito especial ao Prof^o Fernando Leite, meu orientador e mestre nesses últimos sete anos. Sua colaboração e apoio foram essenciais para o desenvolvimento dessa pesquisa e para a minha formação acadêmica, profissional e crescimento pessoal. Agradeço à Prof^a Sely Costa, por todos os ensinamentos e inspiração proporcionados desde o início da minha vida acadêmica na Biblioteconomia. Agradeço à Dr^a Patrícia Bertin por todas suas contribuições para a evolução dessa pesquisa. Igualmente, agradeço à Dr^a Rosemeire Tavares, à Dr^a Ana Valéria Mendonça e à Dr^a Marília Freitas por suas valiosas colaborações para o melhoramento do estudo.

Agradeço ao financiamento do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) para o desenvolvimento da pesquisa. Agradeço também à Universidade de Brasília (UnB) por todos os recursos disponibilizados para a realização desse estudo e da minha formação desde a graduação.

Agradeço aos pesquisadores Dr^a Ana Maria Pinto, Dr^a Débora Diniz, Dr^o Fabrício Campos, Dr^o Fernando Melo, Dr^o Jorg Heukelbach, Dr^o José Mauro Peralta, Dr^a Liane de Castro, Dr^o Luciano Cavalcanti, Dr^a Marília Carvalho, Dr^o Pedro Tauil, Raquel Barbosa, Dr^o Roberto Campos, Dr^o Renato Astray, Dr^o Wayner Vieira de Souza e aos outros pesquisadores que não puderem ser nomeados, mas que contribuíram fundamentalmente para os esclarecimentos das questões aqui apresentadas.

Agradeço ao Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT) por todas as oportunidades proporcionadas de atuação e reflexão no mundo da ciência aberta. Nomeadamente agradeço às minhas companheiras Bianca Amaro, Tainá Batista, Danyelle Silva, Raquel Viana, Luísa Café, Maria Fernanda Melis, Rayanne Silva, Washington Ribeiro e Milton Shintaku.

Com amor, agradeço à minha família pelo suporte e carinho que recebi ao longo da vida. Minha mãe, meu pai e minha avó são as raízes e as razões do que sou e do que já realizei. Ao meu companheiro de vida, Tiago Braga, agradeço pelo entusiasmo e estímulo para a conclusão dessa pesquisa. Por fim, agradeço a todas as amigas, que estiveram presente neste processo, ouvindo, palpitando e participando, especialmente à Ale e Fe.

COSTA, M. **Fatores que influenciam a comunicação de dados de pesquisa sobre o vírus da zika, na perspectiva de pesquisadores.** 2017. Tese (Doutorado em Ciência da Informação) - Universidade de Brasília, Brasília, 2017.

RESUMO

O estudo teve por objetivo propor modelo teórico que ilustra os fatores que exercem influência na comunicação dos dados de pesquisa sobre as doenças causadas pelo vírus da zika. A discussão foi fundamentada em perspectivas teóricas da ciência da informação, no contexto teórico em que se insere a discussão a respeito das doenças emergentes, bem como nos princípios relacionados à comunicação dos dados de pesquisa. A pesquisa é de natureza qualitativa e pressuposto filosófico ontológico. Para a coleta e análise de dados foram utilizadas as estratégias relacionadas à Revisão Sistemática da Literatura e da Teoria Fundamentada. Ambas estratégias permitiram alcançar sete categorias teóricas que representam os fatores intervenientes na comunicação dos dados de pesquisa sobre o vírus da zika. A relação entre as categorias, evidenciada pelos dados das entrevistas e pelas categorias preliminares resultaram em uma teoria substantiva acerca do fenômeno. A teoria emergente evidencia a centralidade da expectativa de reconhecimento e recompensas dos pesquisadores, ponderada pelos aspectos relacionados ao financiamento da pesquisa e da cultura acadêmica. Os três fatores exercem influência sobre os macroprocessos da comunicação dos dados de pesquisa, a saber, o uso, a produção e o compartilhamento dos dados. Entre os processos de uso e produção sobressaltaram-se os fatores relacionados a colaboração. Já entres os processos de produção e compartilhamento revelaram-se os fatores relacionados ao tratamento dos dados e ao uso de repositórios de dados de pesquisa. Por fim, o compartilhamento e a possibilidade de reuso dos dados é afetado diretamente pelo contexto social do vírus da zika como uma doença emergente.

Palavras-chave: comunicação científica, dados de pesquisa, ciência aberta, vírus da zika.

COSTA, M. **Factors that influence the communication of research data on the Zika virus, from the perspective of researchers.** 2017. Thesis (PhD in Information Science) - Universidade de Brasília, Brasília, 2017.

ABSTRACT

This study aims to propose a theoretical model about factors that influence the data communication regarding researches related to Zika's diseases. The discussion was based on information science theoretical perspectives, on theoretical context of emerging diseases discussion and on principles related to research data communication. The research is based on qualitative and ontological philosophy assumptions. To collect and analyze data the Literature Systematized Revision and Grounded Theory strategies were used. Both strategies allowed to systematize seven theoretical categories which represent the main factors related to research data communication related to Zika's virus. The categories relation, highlighted by interview data and preliminary categories, resulted in a new theory regarding the phenomenon. The emerging theory highlights how the researchers recognition and reward are central, balanced with aspects related to research funding and academic culture. Between the production and use of the data the collaboration factors were highlighted. When comes to processes between data production and dissemination the data management and data repository use factors were highlighted. Finally, the data dissemination and data reuse are directly affected by the social context of Zika's virus as an emerging disease.

Keywords: scholarly communication, research data, open science.

LISTA DE FIGURAS

Figura 1: Modelo de comunicação científica de Garvey e Griffith.	24
Figura 2: Crescimento dos repositórios de acesso aberto.	31
Figura 3: Crescimento dos periódicos de acesso aberto.	32
Figura 4: Porcentagem dos artigos científicos livremente disponíveis.	32
Figura 5: Categorias da ciência aberta (FOSTER).	35
Figura 6: Categorias da ciência aberta consideradas para a pesquisa.	37
Figura 7: Modelo DDI de ciclo de vida dos dados de pesquisa.	43
Figura 8: Modelo do estudo para o ciclo de vida dos dados de pesquisa.	45
Figura 9: Países que possuem repositórios de dados	47
Figura 10: Tipos de repositórios de dados de pesquisa.	48
Figura 11: Tipos de recursos armazenados nos repositórios de dados de pesquisa	49
Figura 12: Certificados presentes nos repositórios de dados de pesquisa.	50
Figura 13: Softwares dos repositórios de dados de pesquisa.	52
Figura 14: Padrões de metadados presentes nos repositórios de dados de pesquisa.	53
Figura 15: Licenças de distribuição e uso presentes nos repositórios de dados de pesquisa.	55
Figura 16: Linha do tempo da promoção da abertura dos dados de pesquisa.	67
Figura 17: Mapa das declarações.	69
Figura 18: Comparação dos princípios e recomendações discutidos nas declarações analisadas.	71
Figura 19: Dispersão do vírus da zika no mundo entre 2005 e 2015.	74
Figura 20: Dispersão do vírus da zika no Brasil.	75
Figura 21: Dispersão do vírus da zika no mundo.	76
Figura 22: Dispersão do vírus da zika na Europa.	77
Figura 23: Casos confirmados de dengue no Brasil - 2013, 2014 e 2015.	78
Figura 24: Modelo conceitual da pesquisa.	87
Figura 25: Caracterização da pesquisa.	91
Figura 26: Busca na base de dados LISA.	93
Figura 27: Macroprocedimentos metodológicos.	100
Figura 28: Período das publicações analisadas.	104
Figura 29: Origem das publicações analisadas.	105
Figura 30: Origem do primeiro autor dos artigos selecionados.	106
Figura 31: Técnicas de coleta de dados dos artigos selecionados.	107
Figura 32: Tipos de dados de pesquisa e áreas do conhecimento.	111
Figura 33: Coleções de dados em risco.	113
Figura 34: Áreas do conhecimento e as coleções em risco.	114
Figura 35: Modelo de comunicação de dados proposto por Dallmeier-Tiessen et al. (2014).	116
Figura 36: Categorias preliminares e dimensões de análise.	131
Figura 37: Vertentes da categoria “reuso dos dados”.	135
Figura 38: Vertentes da categoria “transparência da ciência”.	136
Figura 39: Vertentes da categoria “visibilidade”.	138
Figura 40: Vertentes da categoria “cultura acadêmica”.	140
Figura 41: Vertentes da categoria “aspectos legais”.	142
Figura 42: Vertentes da categoria “aspectos éticos”.	146
Figura 43: Vertentes da categoria “saúde pública”.	148
Figura 44: Vertentes da categoria “instituição financiadora”.	150
Figura 45: Vertentes da categoria bem público.	151
Figura 46: Vertentes da categoria “custo”.	153
Figura 47: Vertentes da categoria “sistema de informação”.	154
Figura 48: Vertentes da categoria “padrões”.	157
Figura 49: Vertentes da categoria “gestão dos dados de pesquisa”.	159
Figura 50: Vertentes da categoria “política de informação”.	161
Figura 51: Categoria “cultura acadêmica”.	199
Figura 52: Categoria “colaboração”.	204
Figura 53: Categoria “tratamento dos dados”.	213
Figura 54: Categoria “repositórios de dados”.	215
Figura 55: Categoria “reconhecimento e recompensas”.	220
Figura 56: Categoria “doença emergente”.	228
Figura 57: Diagrama das relações entre as categorias teóricas.	234

LISTA DE TABELAS

Tabela 1: Projeção de infecções pelo vírus da zika para 2015.....	74
---	----

LISTA DE QUADROS

Quadro 1: Receita e lucro das editoras comerciais.....	27
Quadro 2: Aumento dos preços de alguns periódicos (em US\$), 1989-2001.....	28
Quadro 3: Princípios da declaração da OECD (2004).	58
Quadro 4: Novos princípios propostos pela OECD (2007).	59
Quadro 5: Condições de abertura dos dados propostos pelo WGDOS.....	60
Quadro 6: Princípios declarados no Panton Principles.....	61
Quadro 7: Licenças recomendadas pelo Panton Principles.....	61
Quadro 8: Princípios para a comunicação dos dados de pesquisa da RCUK.....	62
Quadro 9: Instituições aderentes ao acordo de cooperação internacional.	65
Quadro 10: Princípios e recomendações discutidos nas declarações analisadas.....	70
Quadro 11: Conjunto de recomendações para a comunicação dos dados de pesquisa	72
Quadro 12: Documentos recuperados na base de dados LISA.....	93
Quadro 13: Características da Revisão Sistematizada da Literatura.....	94
Quadro 14: Ficha de análise dos artigos de periódicos.	94
Quadro 15: Procedimentos metodológicos da pesquisa.....	101
Quadro 16: Aceleração da ciência.	133
Quadro 17: Reuso dos dados.	134
Quadro 18: Transparência da ciência.	136
Quadro 19: Assédio aos pesquisadores.	137
Quadro 20: Visibilidade.	137
Quadro 21: Recompensas aos pesquisadores.	139
Quadro 22: Cultura acadêmica.	140
Quadro 23: Aspectos legais.	141
Quadro 24: Aspectos éticos.	145
Quadro 25: Saúde pública.....	147
Quadro 26: Instituição financiadora.....	149
Quadro 27: Bem público.....	151
Quadro 28: Custo.	152
Quadro 29: Sistemas de informação.....	155
Quadro 30: Padrões.....	156
Quadro 31: Gestão dos dados de pesquisa.....	158
Quadro 32: Políticas de informação.....	161
Quadro 33: Categorias preliminares resultantes da RSL.....	163
Quadro 34: Codificação aberta da entrevista - P01.	168
Quadro 35: Codificação aberta da entrevista - P02.	171
Quadro 36: Codificação aberta da entrevista - P03.	173
Quadro 37: Codificação aberta da entrevista -P04.	176
Quadro 38: Codificação aberta da entrevista - P05 e P06.....	180
Quadro 39: Codificação aberta da entrevista - P07	184
Quadro 40: Codificação aberta da entrevista - P08.	187
Quadro 41: Codificação aberta da entrevista - P09.	190
Quadro 42: Codificação aberta da entrevista - P10.	193
Quadro 43: Categorias abertas.....	195
Quadro 44: Categorias axiais.....	197

LISTA DE SIGLAS

AHRC	Archeology Data Service do Arts and Humanities Research Council
ANPPS	Agenda Nacional de Prioridades de Pesquisa em Saúde
APC	Article Processing Charge
BOAI	Budapest Open Access Initiative
BSD	Berkeley Software Distribution
DACS	Describing Archives: a Content Standard
DANS	Data Archiving and Network Services
DCMI	Dublin Core Metadata Initiative
DDI	Data Documentation Initiative
DIF	Directory Interchange Format
DNA	deoxyribonucleic acid
DOAJ	Directory of Open Access Journals
DSA	Data Seal of Approval
ECDC	European Centre for Disease Prevention and Control
ECDC	European Centre for Disease Prevention and Control
ESRC	UK Data Service do Economic and Social Research Council
FGDC	Federal Geographic Data Committee
FITS	File Information Tool Set
FOIA	Freedom of Information Act
FOSTER	Facilitate Open Science Training For European Research
GFDL	Free Documentation License
GPL	GNU General Public License
H2020	Horizon 2020
ICSU	International Council for Science
ISO	International Organization for Standardization
JISC	Joint Information Systems Committee
KNAW	Academia Real de Arte e Ciência dos Países Baixos
LISA	Library and Information Science Abstracts
MRC	MRC Data Support Service do Medical Research Council
NAM	National Academy of Medicine
NERC	Natural Environment Research Council
NIH	National Institutes of Health
NSF	National Science Foundation
NSF	National Science Foundation
NWO	Organização para Pesquisa Científica dos Países Baixos
OAIS	Open Archival Information System
OECD	Organisation for Economic Co-operation and Development
OKF	Open Knowledge International

OMS	Organização Mundial da Saúde
ONU	Organização das Nações Unidas
OpenDOAR	Directory of Open Access Repositories
OSU	Oregon State University
PGH	Programa Genoma Humano
PIB	Produto Interno Bruto
PNECTIS	Política Nacional de Ciência, Tecnologia e Inovação em Saúde
PREMIS	Preservation Metadata Maintenance Activity
RatSWD	The German Data Forum
RCUK	Research Councils United Kingdom
RSI	Regulamento Sanitário Internacional
RSL	Revisão Sistemática da Literatura
RWA	Research Works Act
SPASE	Space Physics Archive Search and Extract
STFC	Atlas Petabyte Storage do Science and Technology Facilities Council
UNESCO	United Nations Educational, Scientific and Cultural Organization
UNISIST	United Nations International Scientific Information System
USGCRP	U.S. Global Change Research Program
USP	Universidade de São Paulo
WDC	World Data Centre
WDS	World Data System
WGODS	Work Grouping Open Data in Science
XML	eXtensible Markup Language

SUMÁRIO

1	INTRODUÇÃO	14
1.1.	CONTEXTUALIZAÇÃO DO PROBLEMA	14
1.2.	OBJETIVOS	17
1.3.	JUSTIFICATIVA	17
1.3.1.	ASPECTOS TEÓRICOS	18
1.3.2.	ASPECTOS APLICADOS	19
2.	FUNDAMENTAÇÃO TEÓRICA	21
2.1.	A CIÊNCIA E A COMUNICAÇÃO CIENTÍFICA	22
2.1.1.	O MOVIMENTO DE ACESSO ABERTO	28
2.1.2.	A FILOSOFIA ABERTA E A CIÊNCIA ABERTA	33
2.2.	COMUNICAÇÃO DOS DADOS DE PESQUISA	38
2.2.1.	DADOS DE PESQUISA	38
2.2.2.	REPOSITÓRIOS DE DADOS DE PESQUISA	46
2.2.3.	BREVE HISTÓRICO DAS PRINCIPAIS INICIATIVAS RELACIONADAS AOS DADOS DE PESQUISA	56
2.2.4.	PRINCÍPIOS E RECOMENDAÇÕES PARA A COMUNICAÇÃO DOS DADOS DE PESQUISA	68
2.3.	CONTEXTO SOCIAL DA DOENÇA	73
2.4.	CONCLUSÕES DA FUNDAMENTAÇÃO TEÓRICA	83
3.	METODOLOGIA	86
3.1.	MODELO CONCEITUAL DA PESQUISA	86
3.2.	CARACTERIZAÇÃO DA PESQUISA	87
3.3.	PROCEDIMENTOS METODOLÓGICOS	92
4.	APORTE DA LITERATURA SOBRE OS FATORES QUE INFLUENCIAM A COMUNICAÇÃO DOS DADOS DE PESQUISA	103
4.1.	CARACTERIZAÇÃO DA LITERATURA DA CIÊNCIA DA INFORMAÇÃO SOBRE A COMUNICAÇÃO DOS DADOS DE PESQUISA	104
4.1.1.	ATUAÇÃO DE PROFISSIONAIS DA INFORMAÇÃO NA COMUNICAÇÃO DOS DADOS DE PESQUISA	107
4.1.2.	GESTÃO NA COMUNICAÇÃO DOS DADOS DE PESQUISA	110
4.1.3.	POLÍTICAS ABERTAS PARA OS DADOS DE PESQUISA	120
4.1.4.	ÊNFASE NAS CIÊNCIAS DA SAÚDE	125
4.2.	FATORES QUE EXERCEM INFLUÊNCIA NA COMUNICAÇÃO DOS DADOS DE PESQUISA, APRESENTADOS PELA LITERATURA DA CIÊNCIA DA INFORMAÇÃO	130
4.2.1.	CATEGORIAS PRELIMINARES ACERCA DOS IMPERATIVOS DA CIÊNCIA	132
4.2.2.	CATEGORIAS PRELIMINARES ACERCA DO CONTEXTO SOCIAL DA DOENÇA	146
4.2.3.	CATEGORIAS PRELIMINARES ACERCA DO FINANCIAMENTO	148
4.2.4.	CATEGORIAS PRELIMINARES ACERCA DA INFRAESTRUTURA	153
4.3.	CONJUNTO DAS CATEGORIAS PRELIMINARES, SUAS DIMENSÕES E PROPRIEDADES	162
5.	FATORES QUE INFLUENCIAM A COMUNICAÇÃO DOS DADOS DE PESQUISA SOBRE O VÍRUS DA ZIKA NA PERSPECTIVA DOS PESQUISADORES	165
5.1.	DESCRIÇÃO E CODIFICAÇÃO ABERTA	165
5.1.1.	CONSIDERAÇÕES ACERCA DA ENTREVISTA COM O P01	166
5.1.2.	CONSIDERAÇÕES ACERCA DA ENTREVISTA COM P02	168
5.1.3.	CONSIDERAÇÕES ACERCA DA ENTREVISTA COM P03	171
5.1.4.	CONSIDERAÇÕES ACERCA DA ENTREVISTA COM P04	174

5.1.5.	CONSIDERAÇÕES ACERCA DA ENTREVISTA COM P05 E P06	176
5.1.6.	CONSIDERAÇÕES ACERCA DA ENTREVISTA COM P07	181
5.1.7.	CONSIDERAÇÕES ACERCA DA ENTREVISTA COM P08	184
5.1.8.	CONSIDERAÇÕES ACERCA DA ENTREVISTA COM P09	188
5.1.9.	CONSIDERAÇÕES ACERCA DA ENTREVISTA COM P10	190
5.1.10.	A PERCEPÇÃO DE PESQUISADORES A RESPEITO DOS FATORES QUE INFLUENCIAM A COMUNICAÇÃO DOS DADOS DE PESQUISA	194
5.2.	ORDENAÇÃO CONCEITUAL E CODIFICAÇÃO AXIAL	196
5.2.1.	CATEGORIA “CULTURA ACADÊMICA”	198
5.2.2.	CATEGORIA “COLABORAÇÃO”	202
5.2.3.	CATEGORIA “FINANCIAMENTO”	210
5.2.4.	CATEGORIA “TRATAMENTO DOS DADOS”	212
5.2.5.	CATEGORIA “REPOSITÓRIOS DE DADOS”	215
5.2.6.	CATEGORIA “RECONHECIMENTO E RECOMPENSA”	219
5.2.7.	CATEGORIA “DOENÇAS EMERGENTES”	226
5.3.	TEORIA SUBSTANTIVA E CODIFICAÇÃO SELETIVA	233
5.3.1.	EM BUSCA DE RECONHECIMENTO E RECOMPENSA	233
6.	CONSIDERAÇÕES FINAIS	243
<hr/>		
FATORES QUE INFLUENCIAM A COMUNICAÇÃO DOS DADOS DE PESQUISA A PARTIR DA LITERATURA		243
FATORES QUE INFLUENCIAM A COMUNICAÇÃO DOS DADOS DE PESQUISA A PARTIR DA PERCEPÇÃO DOS PESQUISADORES		245
FATORES CONVERGENTES E DIVERGENTES NAS PERSPECTIVAS IDENTIFICADAS SOBRE A COMUNICAÇÃO DE DADOS DE PESQUISA SOBRE AS DOENÇAS CAUSADAS PELO VÍRUS DA ZIKA		246
RAZÕES PELAS QUAIS OS DADOS DE PESQUISA SOBRE O VÍRUS DA ZIKA ESTÃO MAIS PROPENSOS A SEREM COMUNICADOS NO CONTEXTO DA CIÊNCIA ABERTA		247
MODELO TEÓRICO REPRESENTACIONAL DOS FATORES QUE EXERCEM INFLUÊNCIA NA COMUNICAÇÃO DOS DADOS DE PESQUISA SOBRE O VÍRUS DA ZIKA		249
CONTRIBUIÇÕES DO ESTUDO		249
SUGESTÃO PARA NOVOS ESTUDOS:		250
7.	REFERÊNCIAS	252
<hr/>		
APÊNDICES		261
<hr/>		
APÊNDICE A: CODIFICAÇÃO DOS DADOS DA RSL E CATEGORIAS PRELIMINARES		262
APÊNDICE B: ROTEIRO DAS ENTREVISTAS		267

1 INTRODUÇÃO

1.1. Contextualização do problema

A comunicação científica é o conjunto de processos de produção, compartilhamento e uso de conhecimento científico pelos pesquisadores sobre os resultados de suas pesquisas. O conhecimento científico tratado aqui se insere dentro da perspectiva da ciência moderna, que, de acordo com Comte-Sponville (2011), pode ser sistematizada de acordo com quatro características – construção histórica, reconhecimento coletivo, refutabilidade e organização lógica –, as quais dizem respeito à forma pela qual o conhecimento científico é produzido, legitimado, questionado e organizado. Em todas as características dessa ciência moderna, é possível perceber a centralidade da comunicação científica como fator necessário para que ela se realize em condições ideais.

A comunicação científica é um objeto e foi gênese dos estudos da ciência da informação, conforme alegou Borko (1969). Nesse tipo de comunicação, destaca-se o veículo que, há pelo menos cinco séculos, tem sido considerado central para o desenvolvimento da ciência: os periódicos científicos. Segundo Meadows (1999), os periódicos surgiram no século XVII e rapidamente foram se tornando essenciais ao desenvolvimento da ciência.

No século XXI, este veículo tão essencial à ciência passou a ser questionado a partir das possibilidades de comunicação proporcionadas pelas novas tecnologias de informação e comunicação e pelas articulações internacionais em prol do acesso aberto ao conhecimento científico e da ciência aberta. O acesso aberto significa a disponibilização livre e irrestrita dos resultados de pesquisas científicas em texto completo por meio da Internet (LYNCH, 2003). Entre outras abordagens, o acesso aberto tem sido apontado como elemento básico para superar limitações relacionadas tanto ao acesso à literatura científica internacional por países em desenvolvimento, quanto à disseminação do conhecimento científico de pesquisadores de localidades que estão à margem dos periódicos científicos comerciais tradicionais (CHAN et al., 2005).

A perspectiva do acesso aberto tem por base uma “nova filosofia” iniciada nos anos 1990, que foi denominada por Costa (2006) de filosofia aberta. Associadas a essa filosofia, estão as iniciativas dos softwares livres e dos arquivos abertos. A ampliação da noção de abertura e de democratização do acesso trazida por essa filosofia fez emergir um novo modo de pensar, produzir e comunicar a ciência, que recebeu o nome de ciência aberta ou *e-Science*. Nela não são demandadas apenas as publicações científicas para o acesso

público e irrestrito, mas é discutido todo o processo de produção da ciência, desde os softwares utilizados para a pesquisa, em que predominam os softwares livres, até as metodologias de avaliação do alcance da pesquisa, nas quais se discute o uso de métricas alternativas para a ciência. Nesse escopo ampliado, a comunicação dos dados de pesquisa é apontada como uma das condições para a avaliação dos resultados da pesquisa, bem como para o desenvolvimento mais rápido, colaborativo e eficaz da ciência como um todo.

Dados de pesquisa são os dados produzidos ou utilizados para o desenvolvimento de uma pesquisa. Os processos de produção, o compartilhamento e o uso dos dados de pesquisa serão tratados neste estudo como comunicação dos dados de pesquisa. A apropriação da noção de comunicação substancia-se na abordagem de Garvey (1979), que trata dos três macroprocessos (produção, compartilhamento e uso) para elaboração de um modelo de comunicação científica.

Neste estudo, identificou-se que iniciativas para a comunicação dos dados de pesquisa foram desenvolvidas ainda na década de 1950 (RUTTENBERG; RISHBETH, 1994), e, na década de 1970, o tema já é encontrado na literatura da ciência da informação (SØNDERGAARD et al., 2003). Ainda na década de 1990, a discussão se ampliou a partir do Programa Genoma Humano (PGH), que convocou pesquisadores do mundo inteiro para compartilharem seus dados sobre o genoma humano, a fim de contribuir para o seu sequenciamento (ZATZ, 2000).

A partir da experiência do PGH, tornou-se crescente a enunciação de políticas mandatórias ou recomendatórias para a comunicação dos dados de pesquisas, principalmente daqueles resultados de financiamento público. O argumento que se identifica subliminarmente nessas políticas é a noção de que os resultados de pesquisas financiadas com recursos públicos são bens que pertencem a todos, portanto devem estar amplamente acessíveis. No entanto, a perspectiva de publicidade da ciência também é válida para pesquisas que não foram financiadas com recursos públicos, visto que os resultados de pesquisas científicas devem ter amplo acesso a fim de garantir as características da ciência moderna, como a construção histórica do conhecimento, a avaliação e a refutação (BOULTON, 2013).

O pressuposto da ciência produzida com recursos públicos é que ela deve atender aos anseios de toda a sociedade e contribuir para o desenvolvimento social. Duas marcantes iniciativas para a comunicação de dados evidenciaram esses propósitos. Uma delas, citada anteriormente, foi o PGH, que, ao convocar os pesquisadores a compartilharem os dados do sequenciamento do genoma humano, mostrou a importância dessa ação para a cura de doenças que afetam grande parte da população mundial, como

a aids e a tuberculose. A outra iniciativa foi a política de comunicação de dados de pesquisa do National Institutes of Health (NIH) dos Estados Unidos da América, que tem por finalidade otimizar o desenvolvimento científico na área das ciências da saúde e, assim, contribuir para o melhoramento da saúde humana. As duas propostas alegam que a comunicação dos dados auxiliará no melhoramento da saúde coletiva.

As articulações internacionais em prol da ciência aberta e do acesso aberto têm defendido que o amplo acesso às informações científicas, tais como publicações e dados de pesquisa, é capaz de otimizar os esforços científicos e acelerar o avanço da ciência. Uma iniciativa internacional recente parece confirmar essa expectativa. Em janeiro de 2016, as principais instituições de pesquisa na área de ciências da saúde e notáveis editores científicos da área assinaram uma declaração na qual se comprometeram a disponibilizar livremente a comunicação científica, no formato de dados e de publicações, sobre o vírus da zika e o ebola (DYE et al., 2016).

O vírus da zika é conhecido desde 1947 no continente africano. No entanto, segundo Vasconcelos (2015), apenas em 2004 foi registrada uma epidemia em decorrência do vírus em países africanos e asiáticos. Em 2015, efetuaram-se os primeiros registros do vírus no Brasil, onde, até março de 2016, haviam sido registrados 500 casos de suspeita de zika, com três mortes confirmadas (BRASIL, 2015b). No mesmo ano, o European Centre for Disease Prevention and Control (ECDC) registrou alguns casos de pessoas infectadas pelo vírus em países europeus (WHO, 2016b), coincidindo com o período de grande repercussão internacional da temática.

O vírus da zika, além de ser transmitido pelo mesmo vetor que transmite a chikungunya, a dengue e a febre amarela, possui, em relação a essas doenças, semelhanças nos sintomas, nos modos de prevenção e de tratamento (CHAVES et al., 2015). O que chama a atenção é que, muito antes da incidência do vírus da zika, a dengue já era uma doença conhecida no Brasil e em outros países em desenvolvimento. Só no Brasil, 1,5 milhão de casos confirmados foram registrados e houve 1.900 óbitos em decorrência da dengue nos últimos 3 anos (BRASIL, 2016a). Entretanto, somente com a incidência do vírus da zika em algumas regiões do planeta, o tema se tornou uma preocupação internacional, com ações de comunicação de dados e de publicações no contexto da ciência aberta, objetivando a promoção de um esforço comum na superação da doença. Portanto, impõe-se como relevante analisar quais foram as condições que levaram o vírus da zika a tornar-se uma preocupação internacional e um objeto de colaboração científica aberta.

A partir do cenário delineado, o estudo buscou responder às seguintes questões: considerando a percepção de pesquisadores, quais são os fatores que exercem influência na comunicação dos dados de pesquisa sobre as doenças causadas pelo vírus da zika? Quais são as razões pelas quais dados de pesquisas sobre as doenças causadas vírus da zika estão mais propensos a ser comunicados no contexto da ciência aberta do que outras doenças?

1.2. Objetivos

Considerando o contexto exposto, o presente estudo tem como objetivo geral *propor modelo teórico que represente os fatores que exercem influência na comunicação dos dados de pesquisa sobre o vírus da zika, sob a ótica de pesquisadores das ciências biológicas e as ciências da saúde*

Para tanto, a pesquisa deverá contemplar os seguintes objetivos específicos:

- 1) caracterizar, a partir da literatura científica, fatores que influenciam a comunicação de dados de pesquisa;
- 2) identificar, por meio de entrevistas, a percepção de pesquisadores a respeito dos fatores que influenciam a comunicação de dados de pesquisa;
- 3) analisar os fatores convergentes e divergentes nas perspectivas identificadas sobre a comunicação de dados de pesquisa sobre as doenças causadas pelo vírus da zika;
- 4) examinar as razões pelas quais dados de pesquisas sobre as doenças causadas pelo vírus da zika estão mais propensos a ser comunicados no contexto da ciência aberta do que outras.

1.3. Justificativa

A justificativa para o desenvolvimento deste trabalho será apresentada sob dois ângulos. O primeiro diz respeito aos aspectos teóricos da discussão, nos quais se inserem os temas da comunicação científica, incluindo os dados de pesquisa e os objetivos da ciência aberta. O segundo aspecto trata das contribuições práticas da pesquisa no contexto em que ela se insere, ou seja, na comunicação de conhecimento científico sobre as doenças causadas pelo vírus da zika.

1.3.1. Aspectos teóricos

A discussão proposta pelo estudo trata de um fenômeno que tem ocorrido dentro das práticas de comunicação científica – a produção, o compartilhamento e a reutilização de dados de pesquisa. Os processos relacionados à comunicação científica são objetos relevantes de estudo na ciência da informação. Segundo Borko (1969), a ciência da informação tem sua gênese nas preocupações com os processos que envolvem a informação científica, e a sua criação se justifica pela necessidade dos institutos de pesquisa de melhorar seus processos de acumulação e transmissão do conhecimento científico. À luz de tais desafios, Borko define a ciência da informação como a disciplina que investiga as propriedades e o comportamento da informação científica, as forças que a governam e os processos para otimizar seu acesso e seu uso.

O fenômeno da comunicação dos dados de pesquisa tem sido identificado como um elemento capaz de melhorar o fluxo de comunicação e otimizar os processos da ciência, conforme argumentam teóricos da ciência aberta. Portanto, faz-se necessário que tal fenômeno seja analisado sob a perspectiva da ciência da informação. Assim, a análise tem potencial para contribuir para as discussões teóricas da área, atualizando-a para um fenômeno contemporâneo e pertinente. Ademais, o corpo de conhecimento da ciência da informação reúne os requisitos necessários para colaborar com as discussões propostas pela ciência aberta e pelo acesso aberto, no intuito de melhorar as práticas de comunicação dos dados de pesquisa.

O tema dos dados de pesquisa no âmbito da comunicação científica para a ciência da informação retoma uma discussão iniciada na década de 1970 e que ainda não foi sistematicamente desenvolvida. Embora o tema tenha sido introduzido na ciência da informação há mais de quatro décadas, ainda são poucos os estudos que tratam dos dados de pesquisa no âmbito da comunicação científica, conforme constatou este estudo¹. A abordagem aqui proposta busca preencher a lacuna de conhecimento sobre o tema, considerando a perspectiva dos pesquisadores envolvidos no processo, aspecto ainda pouco explorado. Ademais, a noção de comunicação científica proposta por Garvey (1979) é adotada pelo estudo na expectativa de abranger o amplo espectro dos processos envolvidos, conferindo-os a complexidade que lhes são pertinentes.

¹ O tema foi explorado no capítulo 4 da tese.

1.3.2. Aspectos aplicados

As contribuições da pesquisa têm potencial para extrapolar o nível teórico, pois o entendimento a respeito da perspectiva dos pesquisadores sobre o processo e as questões envolvidas na comunicação dos dados de pesquisa é um fator relevante para os conhecimentos acerca da gestão dos dados de pesquisa. Nesse quadro, a pesquisa poderá contribuir fornecendo um diagnóstico capaz de instrumentalizar possíveis intervenções nos processos de comunicação científica, a fim de aprimorá-los.

Uma das questões que se propõe aqui é entender por que outras doenças epidêmicas e que afetam grande parte da população não recebem a atenção devida por parte da ciência. Com isso, espera-se que a análise sobre a comunicação de dados relacionados à pesquisa elaborada a partir da experiência das pesquisas sobre o vírus da zika instrumentalize as mudanças necessárias para outros temas de interesse da saúde pública. A análise sobre os fatores que influenciam a comunicação dos dados sobre o vírus da zika ainda apontou os fatores que favorecem a comunicação de dados de pesquisa e indicou aqueles que inibem a prática. Além disso, os resultados do estudo evidenciaram as razões pelas quais pesquisas sobre algumas doenças estão mais propensas à ciência colaborativa do que outras.

Parte I – Fundamentação teórica e metodologia

2. FUNDAMENTAÇÃO TEÓRICA

O estudo proposto está fundamentado em discussões teóricas da filosofia da ciência, em abordagens da ciência da informação e nos pressupostos do acesso aberto e da ciência aberta. A análise ainda considera o contexto da emergência das doenças causadas pelo vírus da zika e suas aproximações teóricas e sociais com outras doenças, tais como a dengue, a febre amarela e a chikungunya.

Na primeira seção deste capítulo, foram discutidas as características da ciência moderna, a partir da classificação feita pelo filósofo Comte-Sponville (2011). Todas as características apontadas necessitam, em alguma medida, da comunicação científica para que se realizem. Assim, foi traçada uma evolução histórica dos processos de desenvolvimento e comunicação da ciência, a partir da narrativa de Moledo e Olszewicki (2015), que destacam a importância dos dados de pesquisa para a ciência moderna, e de Meadows (1999), teórico da ciência da informação que trata do tema da comunicação científica. Ainda nesta seção, foram analisados dois fenômenos contemporâneos da comunicação científica. Um deles relaciona-se às articulações internacionais em prol do acesso aberto às publicações científicas e o outro trata da filosofia da ciência aberta, que, entre outros objetivos, destaca a relevância da comunicação dos dados de pesquisa para democratização e avanço da ciência.

A segunda seção do capítulo sistematiza as principais iniciativas internacionais de comunicação de dados de pesquisa. A apresentação busca a historicidade das ações e os princípios que as motivaram. Assim, foi possível construir uma linha do tempo dos sistemas de informação, políticas e projetos que promoveram a comunicação dos dados de pesquisa. Também foram discutidos e sistematizados os princípios e as recomendações propostos em cinco documentos que orientam a comunicação dos dados de pesquisa, mostrando suas abrangências, convergências e divergências.

Por fim, discutiu-se o contexto das doenças causadas pelo vírus da zika e suas aproximações teóricas e sociais de outras três doenças: dengue, febre amarela e chikungunya. Essa discussão foi diligenciada a partir dos estudos de Kritski et al. (2007), Teixeira (2006), Vieira e Gomes (2008), Hotez et al. (2006), Zanluca et al. (2015), Morel (2004), Chaves et al. (2015) e dos relatórios produzidos pela OMS (2014) e pelo Brasil (2016), (2015b) e (2014).

2.1. A ciência e a comunicação científica

A ciência é um conjunto de conhecimentos que formam teorias sobre o mesmo tema ou objeto. O filósofo materialista francês Comte-Sponville (2011) atribui a esse tipo de conhecimento quatro características: é construído historicamente, é coletivamente reconhecido, é refutável e é logicamente organizado. A seguir será analisada cada uma das características elencadas.

A construção histórica da ciência se dá com o desenvolvimento dos seus conceitos, paradigmas e estudos. De acordo com Comte-Sponville (2011), o conhecimento científico é muito mais construído do que constatado. Essa afirmativa se opõe à perspectiva defendida por outro filósofo francês, René Descartes, em sua obra publicada em 1637. Descartes (2000) elaborou uma espécie de tratado filosófico no qual defendeu que, a partir do uso da razão, o conhecimento científico poderia demonstrar um fenômeno, ou seja, constatá-lo, e que este era passível de verificação.

O reconhecimento por um grupo específico é uma característica amplamente aceita sobre o conhecimento científico. Toda ciência acontece dentro de uma área que é formada por um grupo de pesquisadores, ou seja, uma comunidade científica. É essa comunidade que validará determinado conhecimento como científico ou não.

A refutabilidade é a característica que considera o conhecimento científico como uma verdade em estado de transição. De acordo com essa perspectiva, defendida por Karl Popper e Thomas Kuhn, toda ciência é construída de paradigmas e teorias, mas essas teorias nunca serão absolutas, por mais que tenham sido aprovadas por uma comunidade científica. Kuhn (2005) defendeu que a transição das teorias é acompanhada pela evolução da ciência. Portanto, as teorias e os paradigmas são datados de uma validade histórica, e, à medida que a ciência se desenvolve, eles podem ser questionados e dar espaço a novos paradigmas e teorias. Nas palavras de Popper (2013, p. 243), a ciência “jamais pode proclamar haver atingido a verdade ou um substituto da verdade”.

A ordenação e a apresentação lógica do conhecimento científico são marcadas pelo pensamento racional dentro da corrente filosófica do racionalismo. Comte-Sponville (2011, p. 500) define o racionalismo como “a doutrina para a qual a razão em nós é independente da experiência [...] e a torna possível”. O filósofo também aponta que essa corrente “se opõe ao irracionalismo, ao obscurantismo e à superstição”. Nesse sentido, a forma de produção de conhecimento na ciência difere, em essência e método, do conhecimento baseado na crença ou na fé.

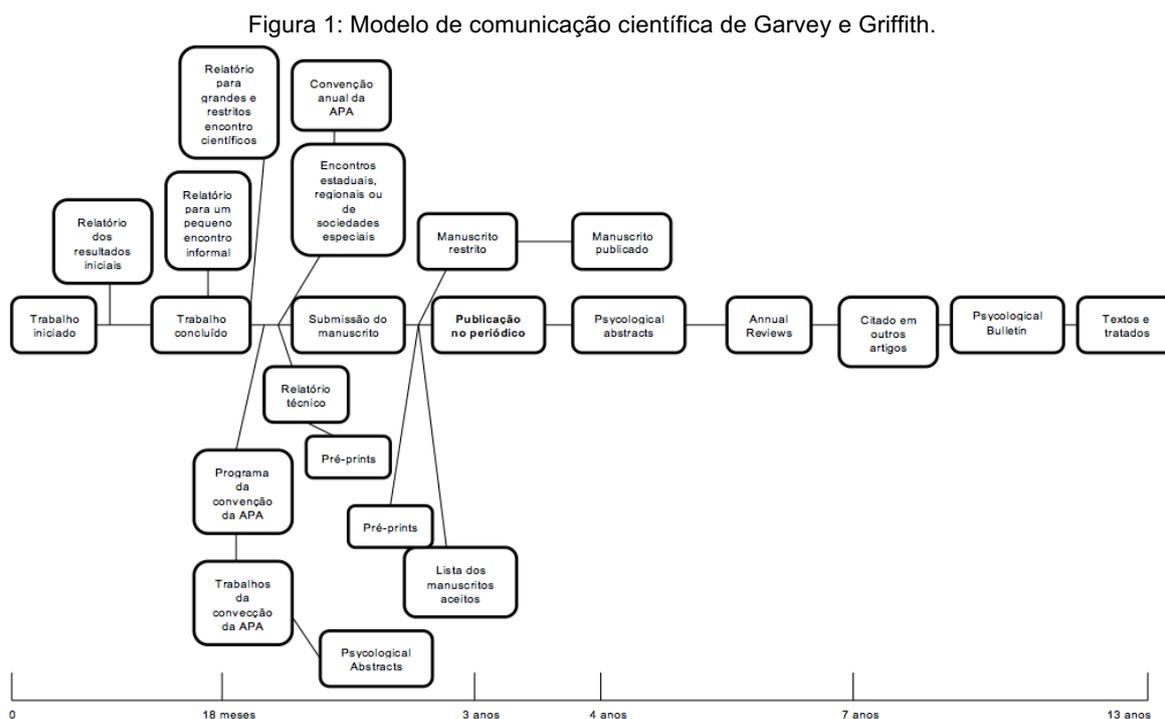
O surgimento do pensamento racional para a produção de conhecimento é atribuído, por Moledo e Olszevicki (2015), a Tales de Mileto. Segundo os autores, a maior contribuição de Tales para o desenvolvimento do pensamento racional foi sua elucidação sobre um eclipse total que aconteceu em 585 a.C. A explicação elaborada por Tales negou o uso de teorias ou argumentos religiosos e mitológicos. Os autores afirmam que a explicação dada por Tales para o eclipse foi baseada em dados observacionais do céu e de seus astros. Segundo a história contada por Moledo e Olszevicki (2015), Tales teria tido contato com tábuas de dados astronômicos de origem mesopotâmica de séculos anteriores ao seu. As tábuas continham descrições minuciosas de observações de constelações, planetas e ciclos astronômicos. A partir desses dados, Tales teria “estabelecido a ideia abstrata de causa (causa natural), que estará presente em toda a ciência posterior” (p. 35).

A gênese da ciência foi marcada pelo pensamento racional, no qual os fenômenos não seriam mais explicados pelas religiões ou mitologias. Além disso, destaca-se o uso de dados observacionais para a explicação de fenômenos naturais. Aqui é oportuno salientar que a explicação racional dada por Tales ao eclipse foi possível mediante seu acesso a dados coletados em épocas anteriores à sua.

O desenvolvimento da ciência e do pensamento racional na Antiguidade resultou na criação de academias e comunidades científicas, demarcadas, principalmente, por limites geográficos. Esses limites evidenciam divergências culturais, barreiras linguísticas e restrição do alcance da comunicação entre os pesquisadores, ou seja, da comunicação científica. Esse tipo de comunicação é essencial para o desenvolvimento da ciência, uma vez que a produção do conhecimento científico toma como ponto de partida os resultados de outros estudos e teorias já existentes em uma área do conhecimento. A comunicação científica é necessária para pelo menos três das quatro características da ciência apontada por Comte-Sponville (2011): a construção histórica, o reconhecimento pelos pares e a refutação.

A comunicação científica é entendida como parte fundamental da própria ciência. De acordo com Roosendaal e Geurts (1998), sua importância relaciona-se aos seguintes aspectos: suas próprias funções de registro da autoria, relacionada com a propriedade intelectual; certificação, que se relaciona com a garantia da qualidade e validação dos conhecimentos gerados; circulação, que promove a acessibilidade aos resultados de pesquisa; e arquivamento, que permite a preservação para uso futuro. Assim, a comunicação científica pode ser entendida como consequência de uma prática de pesquisa, ao mesmo tempo em que é matéria-prima para novos processos de geração de conhecimento.

De acordo com Garvey (1979), para que uma nova ideia seja consideravelmente relevante na ciência, é necessário que ela seja comunicada. Essa comunicação precisa ser compreendida e verificada por outros cientistas para que possa ser considerada válida e ser usada para novas pesquisas, as quais, por sua vez, precisarão igualmente ter seus resultados comunicados. Com isso, segundo o autor, a comunicação torna-se uma característica essencial da própria ciência. Garvey e Griffith propuseram, em 1979, um modelo de comunicação científica que expressava os processos que essa atividade engloba, em relação ao tempo (Figura 1).



Fonte: Garvey, 1979 (com adaptações e traduzido).

O modelo foi construído a partir dos resultados de um estudo sobre a dinâmica dos meios de informação envolvidos nos processos de comunicação e a incorporação dos novos conhecimentos científicos no campo da Psicologia. Uma das motivações dos autores para esse estudo foi a mudança observada em relação à produtividade dos cientistas de Psicologia. De acordo com Garvey e Griffith (1979), antes de 1910, a quantidade de informação produzida na área dobrava a cada 10-12 anos. Depois de 1910, essa quantidade diminuiu, dobrando apenas a cada 15-20 anos. Diante disso, os autores apontaram para duas razões possíveis. A primeira delas é o surgimento de novos tipos de profissionais, dos quais apenas 10% poderiam ser considerados produtivos; e a segunda

é o crescimento do uso da comunicação informal. Em um estudo realizado pelos autores, com 200 pesquisas em Psicologia, observou-se que apenas 15% delas eram originadas de fontes como artigos científicos e anais de eventos.

Embora o modelo seja desatualizado em alguns aspectos e restrito a uma área do conhecimento específico, ele foi utilizado para representar as características e os processos de boa parte das outras áreas. Os processos são referentes a três conjuntos de atividades: produção, compartilhamento e uso da informação.

De acordo com o modelo de Garvey e Griffith (1979), os resultados de uma pesquisa científica em forma de artigo são submetidos a um periódico científico que fará sua avaliação; caso seja aprovado, o trabalho será publicado. Segundo os autores, a publicação transforma em ciência pública a informação científica contida em um manuscrito. Um aspecto que pode ser notado neste ponto da comunicação é a distância temporal entre a conclusão da pesquisa e a publicação em um periódico. De acordo com os dados apresentados por Garvey e Griffith (1979), quando os artigos foram publicados em periódicos, 70% dos autores já haviam iniciado novas pesquisas e, desses, 60% já tinham até concluído um novo trabalho.

Depois de 7 ou 8 meses da publicação do artigo, ele é resumido e publicado no *Psychological Abstracts*. O resumo, então, torna-se uma fonte secundária de informação pública sobre a pesquisa, juntamente com outros trabalhos contemporâneos sobre o mesmo assunto. Os artigos que mostrassem potencial relevância científica reapareceriam no *Annual Reviews of Psychology* 2 ou 3 anos após terem sido publicados.

Os autores ainda calcularam uma faixa de tempo de 2 a 5 anos para que os trabalhos publicados em periódicos e disseminados em outras fontes de informação fossem citados pela comunidade científica. O uso dos resultados da pesquisa por novos estudos e em textos orientadores da disciplina, como tratados, fecha o ciclo da comunicação de um novo conhecimento científico. Segundo os autores, na área de Psicologia, esse ciclo leva um período de 12 a 15 anos.

Considerando as motivações dos autores para o estudo, o modelo de Garvey e Griffith (1979) marca o processo de transferência da informação científica do domínio informal para o formal. Segundo os autores, o fluxo de informação através dos canais informais é relativamente livre de filtros ou monitoramentos e permite que o trabalho atinja um padrão mínimo de cientificidade. Para a transferência para o domínio formal, a informação precisa ser avaliada e aprovada pelos pares do autor, processo que ocorria com a publicação dos artigos em periódicos. Segundo os autores, a fronteira existente entre o domínio formal e o informal ajuda a reduzir o fluxo de informação para os

pesquisadores de uma área, até que a informação seja avaliada. No entanto, os autores ponderam que é necessária a reorganização de alguns dos processos de comunicação científica, de modo a otimizá-los.

A comunicação científica nem sempre teve todos os procedimentos apontados no modelo de Garvey e Griffith (1979), nem esses permanecem do mesmo jeito até hoje, mas os três conjuntos de procedimentos utilizados pelo modelo – produção, compartilhamento e uso – servem de orientação para a análise dos processos de comunicação científica em diversos contextos. Adicionalmente, esse modelo foi base para a proposição de novos modelos de comunicação científica, conforme será discutido nos resultados do segundo objetivo específico deste estudo.

A relevância da comunicação científica para o desenvolvimento da ciência é um aspecto amplamente defendido na literatura pertinente. De acordo com Meadows (1999), a comunicação científica é o coração da ciência, pelo menos da ciência moderna. Para justificar sua afirmação, o autor discorre sobre as transições que a comunicação científica foi sofrendo no decorrer da história da ciência. Ele inicia o relato com os debates e eventos que aconteciam dentro das academias, nos séculos V e IV a.C. No estágio seguinte, destaca-se a utilização de cartas manuscritas, trocadas entre os próprios pesquisadores, e a publicação² de obras monográficas. Por volta do século XVII, desenvolvem-se as sociedades científicas, que nutrem especial interesse pela comunicação. Conforme relata Meadows (1999, p. 5), os fundadores das sociedades científicas

[...] haviam sido influenciados por Francis Bacon, que no último de seus livros descrevera as atividades possíveis de uma instituição de pesquisa. Segundo uma de suas sugestões, um dos requisitos seria que se concedesse alta prioridade à coleta e análise de informações importantes. Alguns dos membros dessa instituição deveriam passar algum tempo no estrangeiro coletando dados por intermédio de conversas e observações diretas, enquanto outros ficariam na sede, lendo e fazendo resumos da literatura publicada no mundo inteiro.

Complementarmente, foram eleitos pesquisadores do “estrangeiro” para enviar comunicações, por meio de cartas, à sede das sociedades científicas³. A elas couberam as atividades de sistematizar e divulgar os relatos de pesquisas enviados pelos pesquisadores eleitos.

² Neste contexto, o termo publicação refere-se a tornar público o documento.

³ As sedes das primeiras sociedades localizavam-se na Inglaterra e na França.

Surgiram assim, na metade do século XVII, os primeiros periódicos científicos. Meadows (1999, p. 7) atribui três razões para o aparecimento desse meio de comunicação. A principal delas, na perspectiva do autor, é “a necessidade de comunicação, do modo mais eficiente possível, com uma clientela crescente interessada em novas realizações”. Outra razão apontada é a necessidade do debate científico para o avanço da própria ciência. A terceira razão identificada foi a expectativa dos editores em obter lucro com a comercialização dessas publicações.

Além das sociedades científicas, passaram também a publicar periódicos científicos as organizações sem fins lucrativos e as editoras comerciais. De acordo com Crow (2006), as sociedades científicas e organizações sem fins lucrativos publicam quase metade dos periódicos científicos internacionais. No entanto, ele chama a atenção para o declínio dessa proporção nas últimas décadas, ou seja, tem diminuído a quantidade de periódicos publicados por organizações sem expectativa de lucro, enquanto tem aumentado a quantidade de periódicos científicos publicados por editoras comerciais. Segundo o autor, a maior parte dos periódicos publicados com interesses comerciais (cerca de 25.000 em 2005) é de domínio de quatro editoras comerciais, a saber: Elsevier, Springer, Wiley e Informa.plc (conhecida como Taylor & Francis).

Morrison (2012), em sua tese de doutorado, mostrou a lucrativa receita dessas editoras comerciais (Quadro 1). Seus dados⁴ demonstram dois fatos. O primeiro é a alta rentabilidade da comercialização de periódicos científicos. O segundo é que o mercado tem se mostrado em crescimento, apesar das alternativas propostas pelo Movimento de Acesso Aberto.

Quadro 1: Receita e lucro das editoras comerciais.

Editora	Receita	Taxa de lucro	Crescimento da taxa de lucro em relação ao ano anterior
Elsevier (2011)	1,1 bilhão de dólares	36%	Sem informações
Springer's Science + Business Media (2010)	294 milhões de euros (~312 milhões de dólares)	34%	4%
John Wiley & Sons (2011)	253 milhões de dólares	42%	13%
Informa.plc (2011)	145 milhões de libras (~216 milhões de dólares)	32,4%	3%

Fonte: Morrison (2012).

Segundo os dados expostos por Morrison (2012), além das editoras comerciais obterem altas taxas de lucro, entre 32% e 42%, esses valores têm sido crescentes ao longo dos anos, entre 3% e 13%. A lucratividade da comercialização da comunicação científica

⁴ Segundo Morrison (2012), os dados foram extraídos da revista *Economist*, número 25 de 2011.

foi um ponto fortemente questionado no início dos anos 2000, nas articulações do Movimento de Acesso Aberto, que propõe uma alternativa a esse modelo.

2.1.1. O Movimento de Acesso Aberto

De acordo com Lynch (2003), o acesso aberto significa a disponibilização livre e irrestrita dos resultados das pesquisas científicas em texto completo, por meio da Internet. Esse objetivo e algumas iniciativas foram se fortalecendo com a constituição do Movimento de Acesso Aberto. O movimento é uma articulação internacional de pesquisadores e outros atores da comunicação científica, que, insatisfeitos com o modelo tradicional de comunicação por meio de periódicos de editoras comerciais, propuseram vias alternativas para a comunicação científica.

A insatisfação da comunidade acadêmica é decorrente da crise dos periódicos. A crise representou a incapacidade de bibliotecas do mundo inteiro em manter o acesso aos periódicos em razão do aumento no valor das assinaturas dos principais periódicos internacionais (VAN DE SOMPEL; LAGOZE, 2000). Soares (2011, p. 13) reúne dados que demonstram a elevação no valor das assinaturas (Quadro 2). O autor alega que o aumento não é recente, mas “acelerou nas últimas décadas – tivemos aumentos de mais de 1 mil por cento entre 1989 e 2001”.

Quadro 2. Aumento dos preços de alguns periódicos (em US\$), 1989-2001.

Título da revista	Preço em 1989	Preço em 2001	Aumento percentual
American Journal of Physical Anthropology	540	1.490	176%
Chemical Engineering Science	650	4.306	562%
Clinical Psychology Review	75	984	1.212%
Communications on Pure and Applied Mathematics	280	2.190	682%
Developmental Biology	936	4.715	404%
Energy Conversion and Management	305	3.019	889%
IEEE Proceedings: All-Society Package	5.335	29.995	462%
Journal of Physics A	1.124	6.030	436%
Educational Research Quarterly	275	505	184% (em um ano)
	(2000)		

Fonte: Soares (2004, p. 14).

Além da crise dos periódicos, o advento da internet e dos computadores pessoais impulsionaram as iniciativas de acesso aberto. A partir desse contexto, o Movimento de Acesso Aberto propôs duas estratégias para a disponibilização livre e irrestrita da literatura científica. Uma delas, a via dourada, corresponde a publicações de artigos em periódicos científicos de acesso aberto, os quais devem disponibilizar gratuitamente por meio da

Internet todos os artigos que publicam. A outra estratégia, a via verde, diz respeito ao depósito dos artigos em repositórios de acesso aberto. Nos repositórios, devem ser armazenados e amplamente disponibilizados os artigos publicados pelos periódicos de acesso aberto e pelos periódicos de acesso restrito, bem como outras publicações científicas.

Ambas as estratégias foram inicialmente definidas na declaração intitulada *Budapest Open Access Initiative* (BOAI), que se estabeleceu como documento central para as discussões sobre acesso aberto (SARMENTO E SOUZA et al., 2005; COSTA, 2014). O documento foi publicado em 2002, como resultado de uma reunião ocorrida em Budapeste no ano de 2001. Na declaração, também é levantada uma discussão sobre a problemática da comunicação científica no contexto contemporâneo; são apontadas experiências já existentes de alternativas para a comunicação científica; e são listadas as vantagens do acesso aberto, quais sejam: aumento da visibilidade, do número de leitores e do impacto da publicação. Esses argumentos são utilizados recorrentemente para demonstrar os benefícios do acesso aberto, embora ainda sejam objeto de controvérsia (COSTA, 2014).

Em 2012, a BOAI completou 10 anos e publicou um novo documento intitulado *Ten years on from the BOAI*. Nessa edição, são reafirmadas as estratégias propostas pela BOAI: repositórios digitais e periódicos de acesso aberto. Adicionalmente, são trazidas novas recomendações para os próximos 10 anos, referentes ao uso de políticas de acesso aberto pelas instituições produtoras de conhecimento científico, ao licenciamento do conteúdo em acesso aberto, à sustentabilidade e coordenação das iniciativas e, por fim, à promoção dos sistemas de acesso aberto.

No mesmo ano em que a BOAI completa 10 anos, estava em discussão uma medida do governo britânico que foi vista como um retrocesso para os objetivos do acesso aberto. A medida adotada foi a modificação da política de acesso aberto destinada às pesquisas que recebem financiamento público no país. A alteração foi no sentido de privilegiar o acesso aberto híbrido, por meio da Article Processing Charge (APC), que se baseia no pagamento de uma taxa de publicação como requisito para o artigo ser disponibilizado em acesso aberto. A mudança permite que os autores publiquem suas pesquisas em um periódico de acesso restrito e paguem uma taxa para disponibilizar seu artigo em acesso aberto. No entanto, Babini (2014) ressaltou que os preços cobrados por algumas editoras são inacessíveis para a maioria dos países e projetos de pesquisa existentes. A autora apresentou dados de Bjor e Solomon, publicados em 2012, que mostraram que o valor cobrado pelas editoras comerciais podem chegar até 5 mil libras (cerca de 25 mil reais).

Entre as críticas que tal medida recebeu, destacam-se as preocupações com a distorção dos objetivos do acesso aberto, o desencorajamento de outros países que não terão como arcar com o pagamento da APC e o fortalecimento da mesma lógica comercial da publicação científica tradicional. Diante do contexto retrógrado em relação ao acesso aberto, Babini (2014) questiona por que as instituições, tais como as universidades, deveriam arcar com taxas cobradas pelas editoras. A autora lembra que são essas instituições que financiam a pesquisa, os pesquisadores e, muitas vezes, os próprios avaliadores das editoras comerciais. Ressalta, ainda, que a concessão para exploração econômica do acesso aos resultados das pesquisas que financiam não é vantajosa para a ciência nem para as universidades ou outras instituições da mesma natureza. A comercialização do acesso só é lucrativa para poucas editoras acadêmicas comerciais.

No mesmo sentido, Weitzel (2014, p. 72) alertou que a promoção do acesso aberto via APC “esconde as verdadeiras intenções do forte *lobby* da indústria editorial de periódicos interessada em vender o acesso aberto e obter maiores rendimentos”. O apontamento da autora demonstra-se pertinente, uma vez que esse modelo tem financiado cada vez mais as editoras comerciais de periódicos científicos (HARVIE et al., 2013). Se compararmos a taxa de lucro da maior corporação de periódicos científicos, a Elsevier, é possível perceber que passados mais de 10 anos da articulação do Movimento de Acesso Aberto, essa indústria vem crescendo e se fortalecendo. De acordo com Morrison (2012), a taxa de lucro da Elsevier, em 2011, foi de 36%; já em 2013, Morrison (2015) assegurou que esse valor subiu para 39%.

De acordo com Taylor (2012), isso significa que:

Quando você paga \$ 37,95 para baixar um PDF de uma revista Elsevier, \$ 13,56 vai direto para os bolsos dos acionistas Elsevier.

Quando você paga \$ 3.000 para ter sua submissão a um periódico Elsevier disponível em acesso aberto, \$ 1.072,20 vai direto para os bolsos dos acionistas Elsevier.

Quando sua biblioteca paga US\$ 1,7 milhão por um pacote de assinaturas Elsevier, \$ 607.580 vão direto para os bolsos dos acionistas Elsevier.

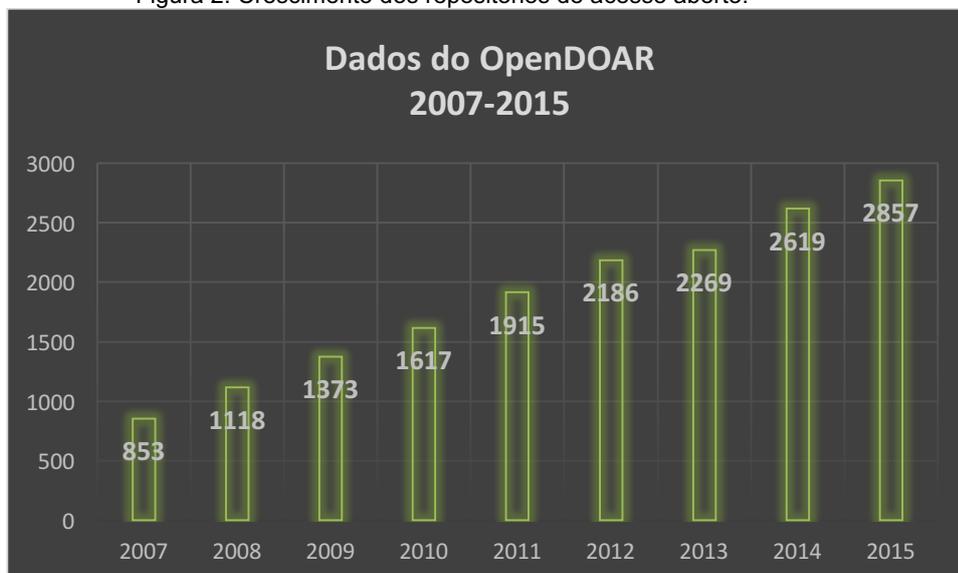
Quando você ou sua biblioteca pagam a Elsevier \$ 23.783 por qualquer razão, é o suficiente para o financiamento da Representante Carolyn Maloney com \$ 8.500 de suborno para co-patrocinar o Research Works Act, fora de seus lucros.

O Research Works Act (RWA), ao qual Taylor se refere, é um projeto de lei norte-americano que foi rechaçado em uma forte movimentação de pesquisadores do mundo inteiro, também em 2012. Segundo Kampf (2012), o “RWA, elaborado em dezembro de 2011, pretendia eliminar a divulgação *online* gratuita de pesquisas financiadas pelo

governo americano, ou seja, com dinheiro público”. Em resposta ao RWA, o pesquisador Timothy Gowers propôs uma ação de boicote à editora Elsevier. No segundo semestre de 2015, a iniciativa já contava com a adesão de mais de 15 mil pesquisadores⁵. Como resultado do boicote, segundo Barros (2012), dez pesquisadores renunciaram ao cargo de editor da Elsevier, o apoio ao projeto de lei foi retirado e um novo projeto de lei de promoção do acesso aberto vem ganhando vigor nos Estados Unidos.

Apesar das forças contrárias ao Movimento, as iniciativas de acesso aberto evoluíram. Morrison (2015) traz alguns dados que nos permitem perceber o crescimento das iniciativas de acesso aberto em um contexto maior. A autora utiliza-se dos dados do Directory of Open Access Repositories (OpenDOAR) para demonstrar que já foram criados quase 3 mil repositórios e disponibilizados cerca de 71 milhões de documentos nesses sistemas (Figura 2).

Figura 2: Crescimento dos repositórios de acesso aberto.

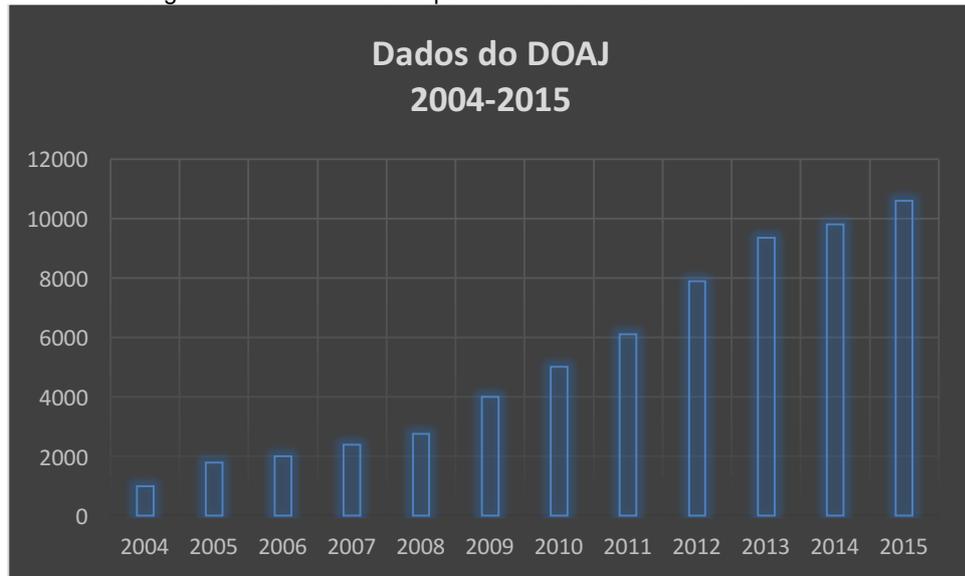


Fonte: Morrison (2015) traduzido e com adaptações.

Morrison (2015) também revelou dados sobre os periódicos científicos, coletados por meio do Directory of Open Access Journals (DOAJ). De acordo com os dados, no início de 2015 já existiam mais de 10 mil periódicos de acesso aberto cadastrados no diretório e cerca de 1,8 milhão de artigos publicados por esses periódicos (Figura 3).

⁵ Dados do site *The cost of knowledge*, onde os pesquisadores aderem ao boicote. Fonte: <http://thecostofknowledge.com/>.

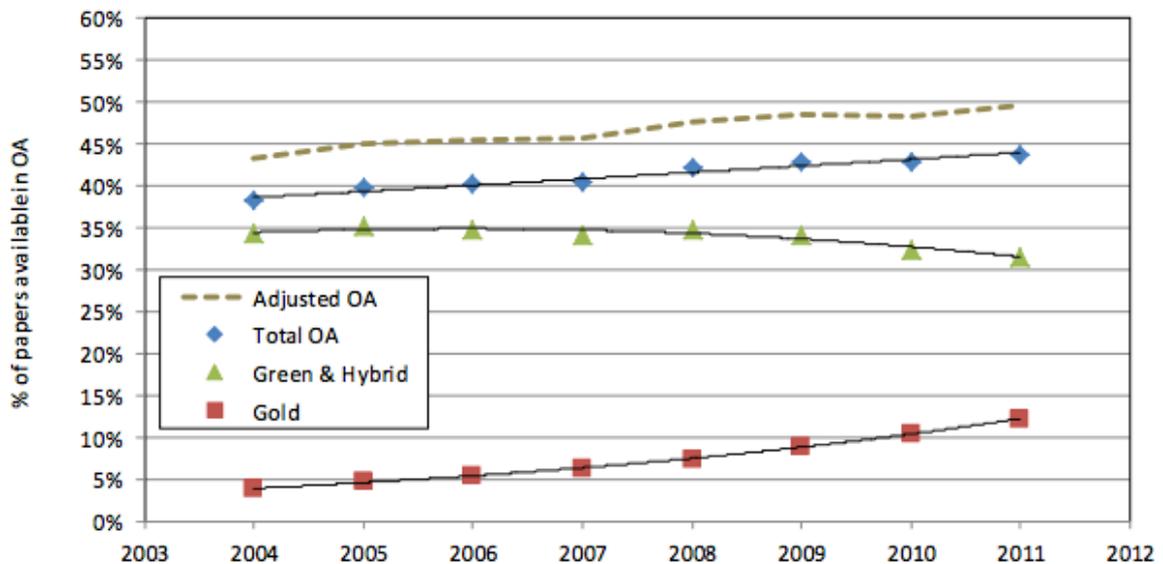
Figura 3: Crescimento dos periódicos de acesso aberto.



Fonte: Morrison (2015) traduzido e com adaptações.

Apesar dos expressivos dados de crescimento das iniciativas de acesso aberto, o percurso para alcançar seus objetivos ainda parece longo. Conforme o relatório elaborado por Archambault et al. (2013), faltam ainda cerca de 50% dos artigos científicos a serem disponibilizados em acesso aberto (Figura 4).

Figura 4: Porcentagem dos artigos científicos livremente disponíveis.



Fonte: Archambault et al. (2013, p. 13).

Com base nesse breve histórico do Movimento de Acesso Aberto, é possível perceber que sua trajetória não está apenas marcada por retrocessos, mas também por avanços, principalmente em relação ao início da adesão dos pesquisadores. Em termos básicos, as proposições e iniciativas do acesso aberto pertencem a um conjunto de ideias que Costa (2006) denominou de filosofia aberta. Na opinião da autora, essas ideias “denotam um novo modelo de representar um igualmente novo processo de comunicação científica, ao mesmo tempo em que serve de base para interpretá-lo”.

2.1.2. A filosofia aberta e a ciência aberta

A filosofia aberta representa uma nova forma de se relacionar e de gerar conhecimentos e produtos de conhecimento. Segundo Costa (2006), nos últimos anos tem sido observado um movimento de criação e uso de ferramentas, estratégias e metodologias que, entre outras coisas, tem alterado o modo de se produzir ciência. Além do campo da ciência, a autora aponta as experiências relacionadas ao desenvolvimento de softwares livres e o compartilhamento de arquivos abertos como resultantes dessa filosofia.

No campo científico, ressaltam-se as discussões sobre a ciência aberta, provocadas pelos desdobramentos da filosofia aberta. Segundo Boulton (2013), os princípios que sustentam a ideia da ciência aberta são os mesmos que serviram de base para as revoluções científicas dos séculos XVIII e XIX, visto que, em ambos os contextos, a motivação seria tornar a ciência pública. O autor cita como gênese dessas ideias o pioneirismo de Henry Oldenburg, que solicitou que suas cartas, enviadas à sede da associação científica Royal Society, fossem publicadas com o objetivo de atingir volumoso número de interessados sobre o tema. Oldenburg também solicitou que a publicação das cartas fosse feita em sua língua vernácula, e que suas evidências fossem publicadas para comprovar suas teorias. Nesse caso, entendem-se as evidências como os dados coletados pela pesquisa e utilizados para sustentação teórica de sua argumentação. Conforme o autor, a prática de tornar públicas as evidências que sustentam uma teoria favorece o princípio científico da refutação e do que ele chamou de *scientific self-correction*⁶.

Os dados de pesquisa também são elementos essenciais para o cumprimento dos quatro objetivos da ciência aberta apontados por Gezelter (2009). O primeiro deles trata da transparência da metodologia, da observação e da coleta dos dados. O segundo objetivo pontua a disponibilidade pública dos dados de pesquisa e a permissão para sua

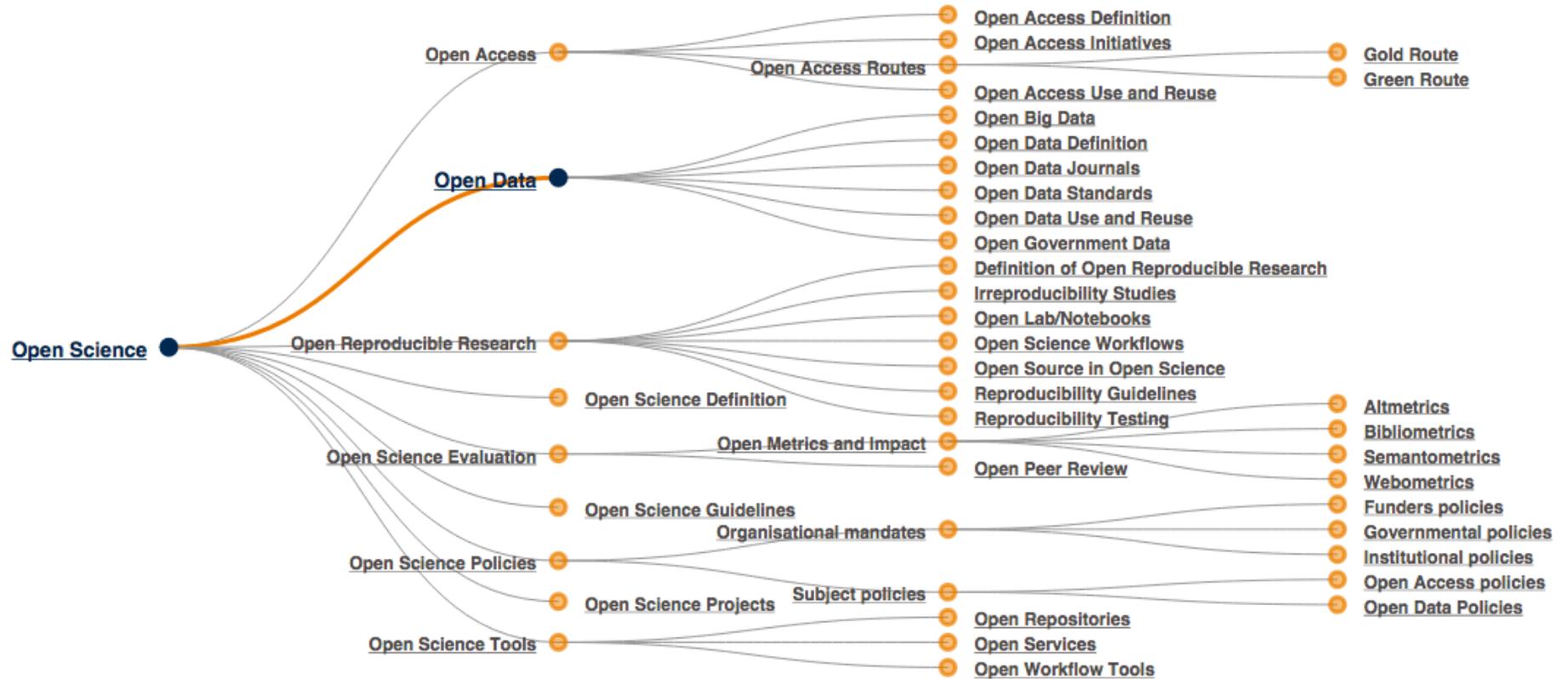
⁶ Capacidade de auto-correção da ciência.

reutilização. O terceiro diz respeito à disponibilização de forma aberta das publicações científicas. O quarto objetivo evidencia a necessidade do uso de ferramentas baseadas na web com vistas a facilitar a colaboração científica. De acordo com o autor, tais objetivos representam a essência dos projetos relacionados ao acesso aberto, aos dados abertos e aos softwares livres.

O que foi pontuado por Gezelter (2009) como objetivos fundamentais da ciência aberta está diretamente relacionado com o que Lopez (2015) denominou de cinco princípios da ciência aberta, que inclui os seguintes elementos: metodologia aberta, software e hardware livres, dados abertos, acesso aberto e avaliação aberta. Segundo esses princípios, a produção da ciência aberta requer que seja utilizada uma metodologia aberta, na qual sejam empregados métodos, processos e documentos relevantes. As tecnologias envolvidas, tanto os softwares quanto os hardwares, devem ser de código-fonte aberto, ou seja, livres. Os dados obtidos na pesquisa devem estar livres para acesso e reutilização. As publicações resultantes da pesquisa devem estar disponíveis gratuitamente na Internet e com o mínimo possível de restrições. Por fim, a ciência aberta poderá ser avaliada a partir de novos critérios e perspectivas, que incluem transparência nos critérios de avaliação e de qualidade, bem como o uso de métricas alternativas. Todas as características serão exploradas a seguir.

No mesmo sentido discutido por Lopez (2015), o projeto Facilitate Open Science Training For European Research (FOSTER) indica seis categorias para se pensar a ciência aberta (Figura 5). A primeira delas trata especificamente das iniciativas do Movimento de Acesso Aberto e suas estratégias: a via verde e a via dourada. Nessa categoria, considera-se que a publicação científica deve estar disponível em acesso aberto, seja por meio dos repositórios, seja pelos periódicos científicos. A segunda categoria trata dos dados abertos, que incluem ampla natureza de dados, desde dados de pesquisa até dados de governo. No que tange aos dados de pesquisa, eles devem acompanhar a publicação, citada na primeira categoria. A terceira trata de aspectos sobre a reprodutibilidade da pesquisa. A quarta categoria diz respeito às formas de avaliação da ciência, que devem considerar métricas abertas e que qualifiquem o seu impacto em sentido amplo. No documento são citados quatro exemplos das novas formas de avaliação, quais sejam: métricas alternativas, bibliometria, semantometria, webmetria. A quinta categoria ocupa-se das políticas da ciência aberta. Por fim, a última categoria trata das ferramentas que serão utilizadas para pesquisa, que devem ser preferencialmente softwares e hardwares livres.

Figura 5: Categorias da ciência aberta (FOSTER).



Fonte: <https://www.fosteropenscience.eu/foster-taxonomy/open-science>

Nesse âmbito, observa-se a aproximação das categorias utilizadas por Lopez (2015) e FOSTER (2015). Ambas possuem cinco categorias em comum, e o projeto FOSTER propõe uma categoria a mais (Figura 6). As categorias comuns são as seguintes:

- a) Acesso aberto às publicações científicas – disponibilização do artigo científico na Internet, sem custos e restrições para o acesso. O acesso aberto pode ser proporcionado pelo próprio periódico ou por uma base de dados interoperável no contexto dos arquivos abertos, ou seja, repositórios digitais.
- b) Dados de pesquisa abertos – os dados coletados e utilizados pela pesquisa devem ser publicados como meio de validar a pesquisa comunicada, permitir a refutação do argumento proposto e promover o desenvolvimento do tema estudado. Sua publicação deve ser feita em uma base de dados específica para a área do conhecimento ou segundo as orientações dos editores científicos. O conjunto de dados deverá receber um identificador persistente, que será relacionado ao artigo do periódico.
- c) Metodologias abertas – a pesquisa deve ser desenvolvida a partir de métodos e processos consistentes. Isso permitirá que ela possa ser reproduzida ou analisada corretamente.
- d) Avaliação aberta – tanto o artigo do periódico quanto os outros produtos da pesquisa deverão considerar avaliações alternativas, além das tradicionais. Nas novas formas de avaliação, destacam-se as métricas alternativas e a webmetria.
- e) Ferramentas abertas – a pesquisa deverá apoiar-se em tecnologias abertas, tanto softwares como hardwares livres.

Além das cinco categorias comuns às duas propostas, considerou-se que a sexta categoria sugerida por FOSTER (2015) é relevante para se pensar a ciência aberta.

- f) Políticas abertas – são os documentos das instituições ou agências fomentadoras de pesquisa que promovem as iniciativas de acesso aberto nas pesquisas que financiam.

Figura 6: Categorias da ciência aberta consideradas para a pesquisa.



Fonte: Elaboração própria a partir das propostas de Foster (2015) e Lopez (2015).

As categorias expostas não são critérios para julgar a produção de conhecimento a partir dos princípios da ciência aberta. Elas funcionam como dimensões de análise da prática de pesquisa e dos valores amplamente debatidos e relevantes para tornar os resultados da ciência acessíveis a todos.

A ciência aberta também tem sido discutida sob a perspectiva da *e-Science*, que, segundo Hey e Trefthen (2008), representa uma nova geração de problemas científicos, uso de ferramentas e tecnologias para construção do conhecimento de modo colaborativo. Os autores observam que essa nova geração de problemas alcança questões que vão desde a simulação de engenharia complexa até o desenvolvimento de fármacos. A mesma perspectiva é defendida por Medeiros e Caregnato (2012), ao observarem que a “*e-Science* altera fundamentalmente a maneira com que os cientistas realizam seu trabalho, as

ferramentas que usam, os tipos de problemas que abordam e a natureza da documentação e da publicação que resulta da sua pesquisa” (p. 315).

A expectativa da *e-Science* é construir uma rede de trabalho global, conforme perceberam Treleaven e Emmott (2006). Segunda tal perspectiva, a articulação da rede seria possível com o uso das tecnologias de informação e comunicação já disponíveis. Assim, a ciência moderna, que tem por princípios a universalidade e a publicidade, pode ter seus processos colaborativos intensificados por meio das tecnologias de informação (GRAY, 2009). Atkins et al. (2003) consideram que o conjunto dessas tecnologias, que chamaram de *cyberinfrastructure*, é um elemento de elevação da ciência. Os autores opinaram que a correta disponibilização da *cyberinfrastructure* permitiria que os pesquisadores se ocupassem mais de suas pesquisas e menos de atividades administrativas, tais como a gestão dos dados de pesquisa.

A questão relacionada aos dados parece ser o centro das motivações para a *e-Science*. Para Hey e Trefthen (2008), ela surge como uma tentativa de resposta ao que os autores chamaram de *data deluge* ou *data strom*, conforme denominou Corti et al. (2009). Isso porque, em quase todos os campos do conhecimento, tem sido exponencialmente crescente a quantidade de dados que precisam ser coletados, acessados, armazenados, compartilhados e reutilizados.

2.2. Comunicação dos dados de pesquisa

A comunicação dos dados de pesquisa tem se mostrado relevante nas discussões sobre a ciência aberta. No entanto, segundo Guibault e Wiebe (2013), ainda não existe uma definição comum sobre o conceito “dado de pesquisa”. Por sua vez, Sayão e Sales (2014) destacam que, por vezes, o entendimento sobre dados de pesquisa pode variar entre os pesquisadores e principalmente entre as áreas do conhecimento. Portanto, a revisão de literatura sobre o tema encontrou algumas divergências entre as conceituações de dados de pesquisa, que serão discutidas a seguir.

2.2.1. Dados de pesquisa

Guibault e Wiebe (2013) afirmam que dados de pesquisa são os dados produzidos no decorrer da pesquisa. Já a Organization for Economic Co-operation and Development (OECD,2007) e Padilla Navarro et al. (2013) enfatizam que são os dados utilizados como

fonte primária em uma pesquisa, não necessariamente produzidos pela pesquisa. Rodrigues et al. (2010) ampliam o escopo da definição incluindo todos os dados usados de algum modo na pesquisa. É na mesma noção ampliada de Rodrigues et al. (2010) que o documento norteador para um dos projetos de comunicação de dados de pesquisa na Europa⁷ apresenta o conceito. De acordo com o Guidelines (2013), os dados de pesquisa referem-se à informação, em particular aos fatos ou números coletados para serem examinados e considerados como base para argumentação, discussão ou cálculo.

Considerando as três perspectivas aqui apresentadas, propõe-se uma definição aproximada das propostas para ser utilizada neste estudo. Portanto, entende-se que os dados de pesquisa são os dados produzidos como fontes primárias, ou utilizados para o desenvolvimento de uma pesquisa, os quais têm por finalidade embasar a argumentação da pesquisa.

2.2.1.1. Formas e categorias de dados de pesquisa

Guibault e Wiebe (2013) ressaltam as possíveis formas de apresentação dos dados, como bases de dados brutos, tabelas, gráficos e imagens. Perry (2008) amplia a abrangência para “qualquer informação que possa ser armazenada no formato digital, incluindo texto, números, imagens, áudios ou vídeos, *software*, algoritmos, equações, animações, modelos e simulações”. No documento Guidelines (2013), são exemplificados os tipos de dados abarcados no âmbito do projeto, que inclui todas as formas já citadas e acrescenta os resultados de experimentos, mensurações, observações resultantes do campo de pesquisa, resultados de questionários, resultados de entrevistas, gravações e imagens.

Os dados de pesquisa, independentemente de sua forma, podem ser classificados segundo duas de suas características, conforme propôs o National Science Foundation (NSF) em seu documento intitulado Proposal & Award Policies & Procedures Guide (NATIONAL, 2014). Uma das características diz respeito à natureza dos dados de pesquisa, que podem ser observacionais, computacionais e experimentais. Os dados de pesquisa observacionais são obtidos por observações diretas e por vezes são considerados como registros históricos, por causa de sua função de registro de observações de um momento específico. Os dados de pesquisa computacionais são resultantes de execução de programas computacionais e simulações. Já os experimentais

⁷ O projeto em questão é o Horizon 2020 <[NATIONAL SCIENCE FOUNDATION \(NSF\), 2014](#)> e será apresentado na próxima seção.

são originários de experimentos em situações controladas, portanto é esperado que seja possível reproduzi-los. Além das três categorias propostas pelo NSF, Borgman (2010) ainda acrescenta um quarto tipo de dados que independe da sua natureza. Esses seriam os diversos dados úteis para a pesquisa científica, tais como dados de negócios, de governos e da vida privada.

A segunda característica utilizada para categorizar os dados é o nível de tratamento que os dados recebem. A partir dessa perspectiva, eles podem ser classificados em duas categorias. A primeira é a dos dados brutos (*raw data*) ou formato primário. Esses dados são passíveis de ser submetidos a diversos estágios de processamento e podem ser utilizados por diversas pesquisas, que não participaram de seu processo de produção ou coleta. Essa é a categoria de dados de maior interesse em ser compartilhada pela sua capacidade de ser reutilizada facilmente por outras pesquisas. Faniel e Jacobsen (2010) apontam que o compartilhamento dos dados brutos favorece a economia de recursos para as pesquisas, uma vez que não é necessário recoletar dados já disponíveis. De acordo como os autores, motivados por esses benefícios, agências de fomento dos Estados Unidos e da Inglaterra estariam promovendo sistematicamente o compartilhamento dos dados brutos das pesquisas que financiaram. A segunda categoria é denominada de dados derivados, que são versões de dados brutos após processamento específico (NATIONAL, 2014).

Corti et al. (2014) ainda propõem uma classificação dos dados relacionada ao estágio da pesquisa, em que os dados seriam preliminares ou finais. Os dados preliminares são os dados resultantes dos processos preliminares das pesquisas. Já os dados finais são aqueles que sustentam o argumento dos resultados da pesquisa.

A coleção dos dados de pesquisa (*set data*) é a representação sistemática e parcial do objeto pesquisado, conforme definiu Padilla Navarro et al. (2013). De acordo com os autores, as coleções podem ser classificadas em três categorias. O primeiro tipo de coleção refere-se aos dados coletados a partir das necessidades apenas de uma pesquisa específica e, em razão de sua natureza, é provável que não sejam aplicáveis a padrões existentes. A segunda diz respeito às coleções de dados de uma área do conhecimento e reúne os dados que servem a uma comunidade científica específica. Normalmente são passíveis de ser ou já são objetos de normalização por padrões existentes. É comum que sejam vinculados a agências de fomento ou governamentais, que financiaram em alguma medida sua coleta e/ou organização. A última categoria são as coleções de dados referenciais, que extrapolam os limites de áreas do conhecimento e podem ser utilizados por usuários diversos em formação e finalidade. Normalmente são apresentados em

padrões bem estabelecidos e demandam financiamento para sua preservação em longo prazo (PADILLA NAVARRO et al., 2013).

2.2.1.2. Gestão de dados de pesquisa

A gestão refere-se aos processos de planejamento, manipulação, armazenamento e preservação dos dados de pesquisa. A forma pela qual esses processos serão realizados tende a ser normalizada em contextos específicos, principalmente sob a orientação das agências financiadoras de pesquisa. Ruusalepp (2008) destaca a dificuldade e talvez a impossibilidade de estabelecer um modelo universal de gestão e comunicação dos dados, por causa da relação direta que as ações têm com os termos do financiamento da pesquisa e que podem variar consideravelmente dentro de um mesmo país e certamente em âmbito internacional.

Independentemente do modelo adotado, todas as práticas visam garantir que os dados de pesquisa sejam de alta qualidade, sustentáveis, acessíveis, reutilizáveis e estejam bem organizados, documentados e preservados (CORTI et al., 2014). Considerando as várias dimensões que a gestão dos dados pode assumir, Corti et al. (2014) propõem o uso de um *checklist* baseado em um serviço proposto pela UK Data Service (2013). No documento, são elencadas questões referentes ao planejamento, à documentação, ao formato, ao armazenamento, à confidencialidade, à ética e ao consentimento, ao copyright e ao compartilhamento.

Em todas as discussões acerca da gestão dos dados de pesquisa, o processo do planejamento é apontado como essencial para o conjunto das ações. Corti et al. (2014) salientam que o planejamento influencia em todos os processos que envolvem os dados, até mesmo na forma como eles serão coletados, pois é nesse momento que se identificam as características dos dados, seus limites e os cuidados que eles demandam.

O planejamento para a comunicação dos dados de pesquisa é formalmente descrito em um documento, no qual é esboçada a forma como os dados serão coletados ou gerados e como serão manipulados durante o projeto de pesquisa e os procedimentos necessários após a finalização da pesquisa, incluindo sua preservação. A descrição dos dados coletados inclui as metodologias utilizadas e os padrões pertinentes. Informações referentes à disponibilização pública dos dados também são recorrentemente demandadas, principalmente por agências de fomento à pesquisa que são signatárias de iniciativas de acesso aberto. Nesses contextos, os documentos devem declarar se e como

os dados serão disponibilizados em acesso aberto e quais licenças de uso serão atribuídas a eles.

Grandes agências de fomento aderentes às práticas de comunicação dos dados possuem seus próprios modelos de documentos para o planejamento da gestão dos dados de pesquisa. Além dos modelos, estão disponíveis alguns programas computacionais que visam auxiliar o processo de planejamento. Corti et al. (2014) destacaram o DMPOnline⁸ e o DMPTool⁹, enquanto Ray (2014) sugere o uso do DMVitals Tool¹⁰. Entre as funcionalidades das ferramentas, destaca-se a possibilidade de elaboração do ciclo de vida dos dados.

2.2.1.3. Ciclo de vida dos dados de pesquisa

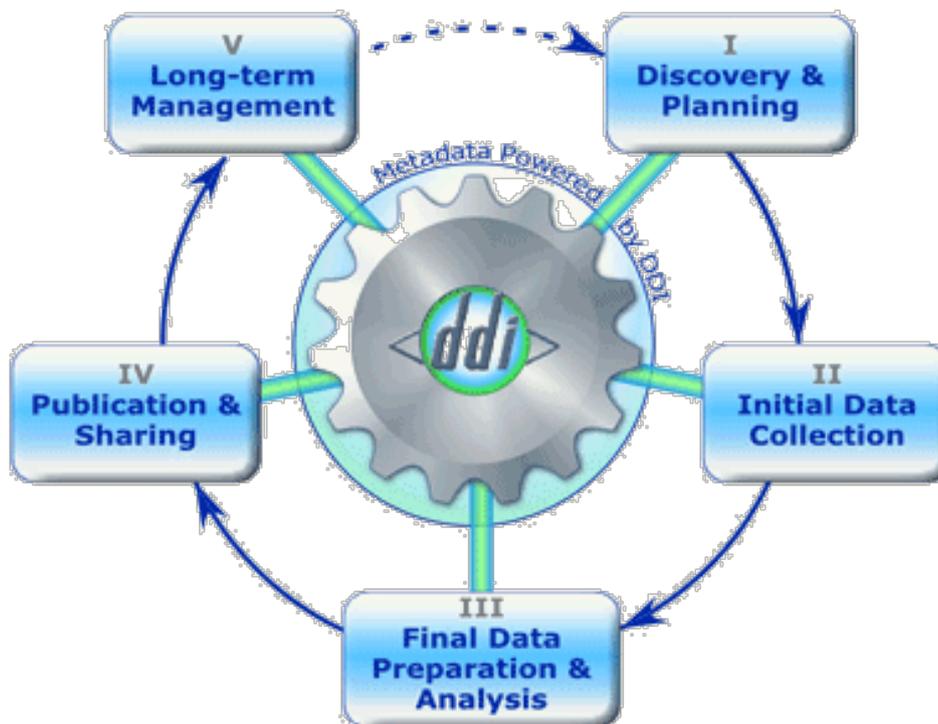
Em 2013, um documento intitulado *Data Documentation Initiative* (DDI) apresentou de forma gráfica e sistematizada os processos relacionados ao planejamento e à curadoria dos dados de pesquisa, que foi denominado de ciclo de vida dos dados de pesquisa. O modelo de ciclo de vida proposto no documento enfatiza cinco processos, os quais são denominados: 1) descoberta e planejamento; 2) coleta inicial; 3) coleta final, preparação e análise; 4) publicação e compartilhamento; 5) gestão de longo prazo (Figura 7).

⁸ Disponível em: <https://dmponline.dcc.ac.uk/>

⁹ Disponível em: <https://dmp.cdlib.org/>

¹⁰ Disponível em: DMVitals Tool

Figura 7: Modelo DDI de ciclo de vida dos dados de pesquisa.



Fonte: Data Documentation Initiative (2013).

Corti et al. (2014) enfatizam a importância do ciclo de vida dos dados de pesquisa, pois, além de ser um instrumento auxiliar para gestão, ele permite que os participantes entendam seus papéis e responsabilidades no contexto geral da comunicação científica. Diante da relevância apontada, os autores propuseram um novo modelo, com base no modelo DDI, ampliando-o. Nesse modelo, foram destacados seis processos, a saber:

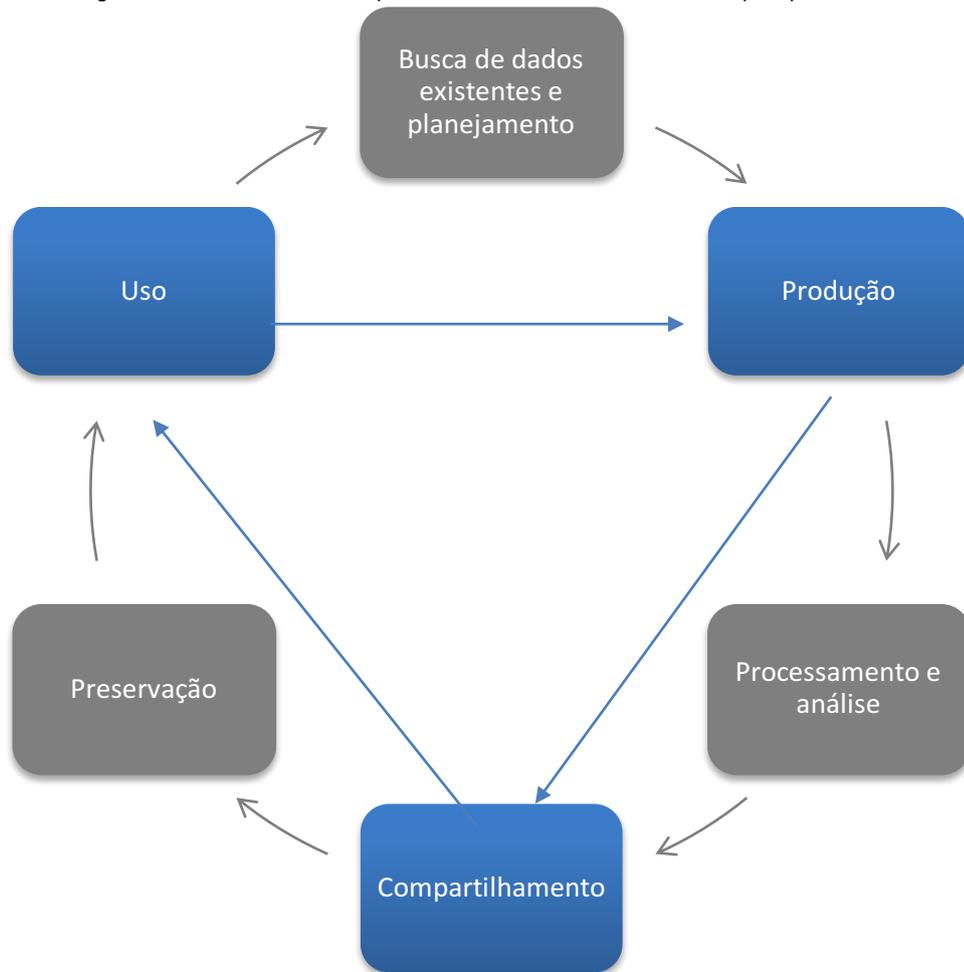
- ✓ Descoberta e planejamento – desenho da pesquisa; planejamento da gestão dos dados; planejamento do consentimento para o compartilhamento; planejamento da coleta; protocolos de processamento e *templates*; busca de fontes de dados já existentes.
- ✓ Coleta de dados – pode ser feita por meio de registro, observação, mensuração, experimentação e simulação; captura e criação de metadados; aquisição de dados de terceiros.
- ✓ Processamento dos dados e análise – inserção de dados, digitalização, transcrição, tradução; checagem, validação, anonimização dos dados quando necessário; derivação dos dados; descrição e documentação;

análise; interpretação; produção de publicações; citação dos recursos dos dados; gestão e armazenamento dos dados.

- ✓ Publicação e compartilhamento – estabelecimento dos direitos autorais dos dados; criação de metadados de descoberta e documentação para o usuário; publicação ou compartilhamento dos dados; controle do acesso aos dados; promoção dos dados.
- ✓ Gestão de longo prazo – migração dos dados para melhores formatos; migração dos dados para meios adequados; backup; compilação e produção de metadados e de documentação; preservação e curadoria dos dados.
- ✓ Reuso dos dados – segunda análise; realização de pesquisas de acompanhamento; condução de pesquisas de revisão; exame minucioso dos resultados; uso dos dados para ensino e aprendizagem.

Nesta pesquisa, a comunicação dos dados de pesquisa considerará o modelo proposto por Corti et al. (2014), que, de forma adaptada, é ilustrado na Figura 8.

Figura 8: Modelo do estudo para o ciclo de vida dos dados de pesquisa.



Fonte: Adaptado da proposta de Corti et al. (2014).

Nele estão presentes os seis processos propostos pelos autores. Além disso, destaca-se a relação do modelo com a perspectiva aqui proposta de comunicação dos dados de pesquisa, em que se ressaltam os macroprocessos de produção, compartilhamento e dados de pesquisa. A relação estabelece que a comunicação é um elemento-chave para a compreensão do ciclo de vida dos dados de pesquisa, ao mesmo tempo que é um instrumento de gestão para a sua comunicação.

Além dos modelos de ciclo de vida, os repositórios de dados de pesquisa são recorrentemente apontados como estratégicos para a efetivação da comunicação dos dados e como instrumento de gestão. Considerando sua relevância para o estudo, o tema será abordado no próximo tópico.

2.2.2. Repositórios de dados de pesquisa¹¹

Os repositórios de dados de pesquisa são bases de dados digitais, onde são armazenados, disseminados e preservados os dados de pesquisa em formato digital. Além dessas funções, Walport e Brest (2011) salientam que eles são uma estratégia eficiente para a organização e preservação dos dados de pesquisa. No entanto, os autores destacam que esses sistemas necessitam estar bem estabelecidos e dispor de padrões e ferramentas que sejam capazes de descrever e divulgar os dados de pesquisa de modo eficiente.

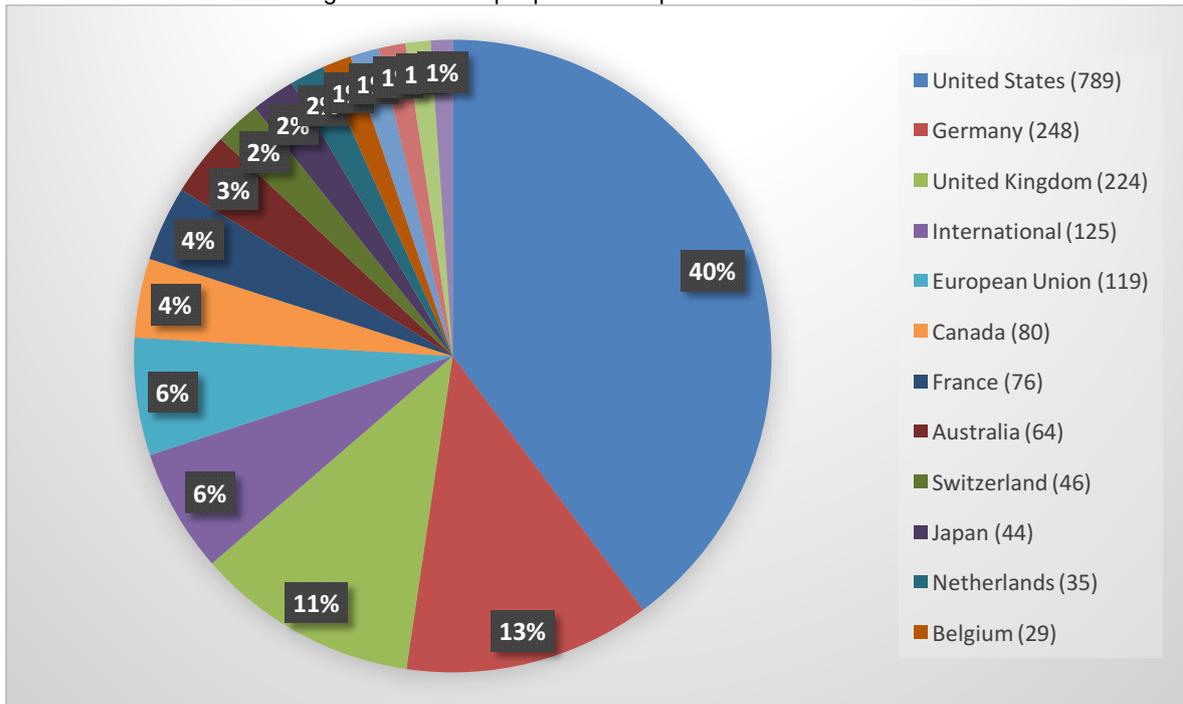
A principal fonte de informação, localizada pelo estudo, sobre os repositórios de dados de pesquisa foi o diretório re3data.org (Registry of Research Data Repositories). O diretório se apresenta como um registro global de repositórios de dados de pesquisa e cobre diversas áreas do conhecimento. Portanto, para traçar um panorama geral dos repositórios de dados no mundo, utilizaram-se os dados disponíveis no re3data.org no primeiro semestre de 2016, período em que haviam 1.558 repositórios registrados. Os resultados da análise foram sistematizados em três categorias: abrangência geográfica e temática, características dos sistemas e promoção da ciência aberta.

2.2.2.1. Abrangência geográfica e temática dos repositórios de dados de pesquisa

Na categoria referente à abrangência, foram analisadas duas variáveis. A primeira delas tratou da expressão geográfica dos repositórios, identificando, assim, os países pelos quais os repositórios estão registrados. O país com maior quantidade de repositórios foram os Estados Unidos da América, de onde se originam 40% dos repositórios de dados do mundo. A Alemanha e o Reino Unido também apresentaram expressividades no universo, representando 13% e 11% respectivamente. Além dos países da América do Norte e da Europa, apenas a Austrália, o Japão e a Índia obtiveram representação entre os países, com pelo menos 1% do conjunto total de repositórios. Com isso, observou-se que nenhum país da América Latina e da África apresentou expressão em quantidade de repositórios de dados de pesquisa (Figura 9).

¹¹ O texto dessa seção foi publicado no seguinte artigo:
COSTA, M.; BRAGA, T. *Repositórios de dados de pesquisa no mundo*. Cadernos BAD, n. 2, p. 80-95, 2016. Disponível em: <http://www.bad.pt/publicacoes/index.php/cadernos/article/view/1585>.

Figura 9: Países que possuem repositórios de dados¹²



Fonte: Elaboração própria

A segunda variável tratou da cobertura temática dos repositórios analisados. Foi identificado que metade (49%) dos repositórios é relacionada às ciências da vida, enquanto a outra metade das áreas de conhecimento divide-se entre ciências naturais (12%), engenharias (12%), ciências sociais e humanidades (26%).

2.2.2.2. Características dos repositórios de dados de pesquisa

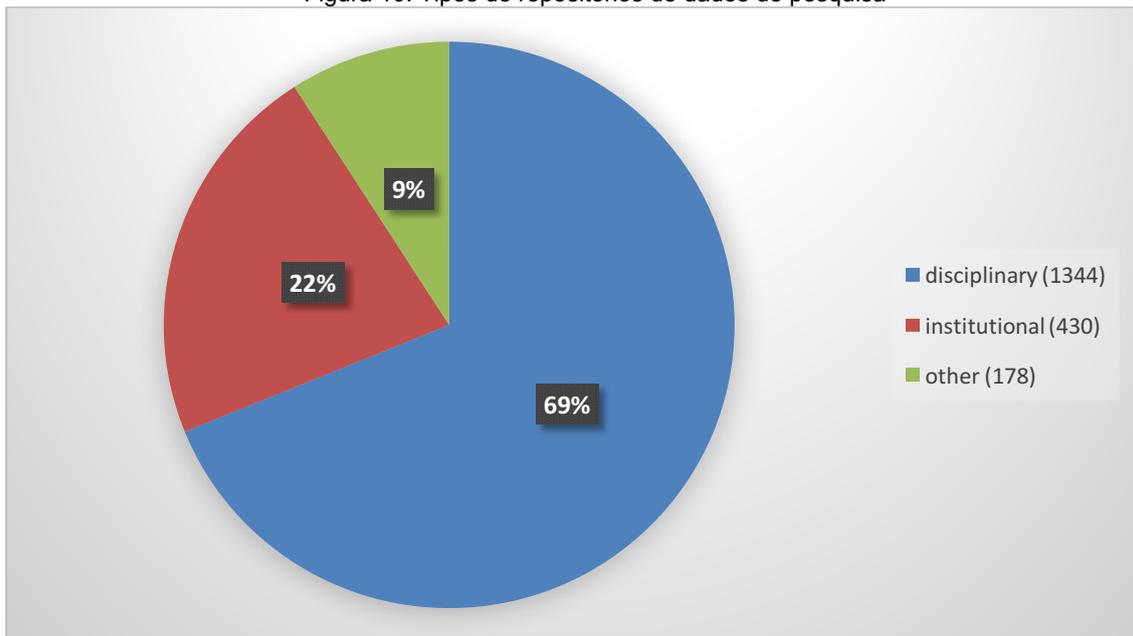
Na segunda categoria de análise, foram exploradas as características dos sistemas: tipo de repositório, tipo de recursos armazenados e certificados aderidos.

Tipos de repositórios

De forma geral, identificou-se que cerca de 70% dos repositórios de dados são temáticos, ou seja, são limitados por área do conhecimento e não por instituições (Figura 10).

¹² No gráfico foram apresentados apenas os países com representatividade maior que 1%.

Figura 10: Tipos de repositórios de dados de pesquisa

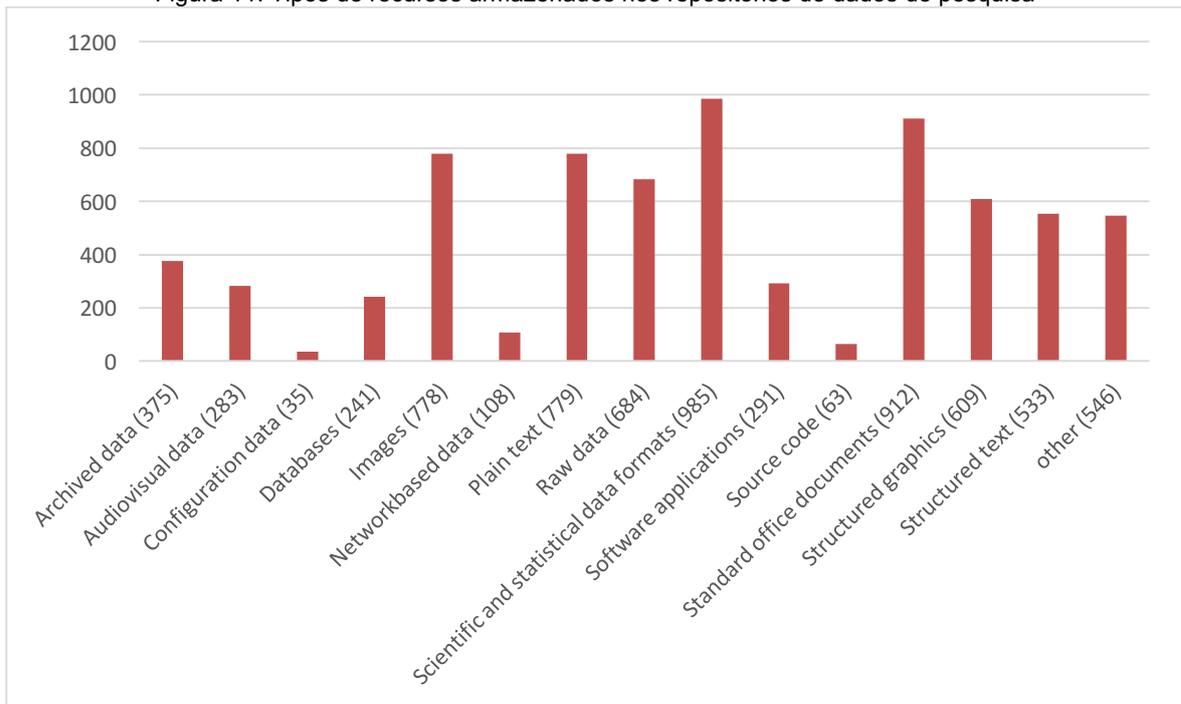


Fonte: Elaboração própria

Tipos de recursos armazenados

Entre os documentos que os sistemas armazenam e disponibilizam, destacam-se quatro tipos: os dados científicos em formatos estatísticos (14%), documentos de texto (office) padrão (13%), imagens (11%) e textos simples (11%). Os outros tipos representam menos de 10% do total de documentos, inclusive os dados brutos, que são recorrentemente citados como necessários para a reutilização dos dados no contexto da ciência colaborativa (Figura 11).

Figura 11: Tipos de recursos armazenados nos repositórios de dados de pesquisa

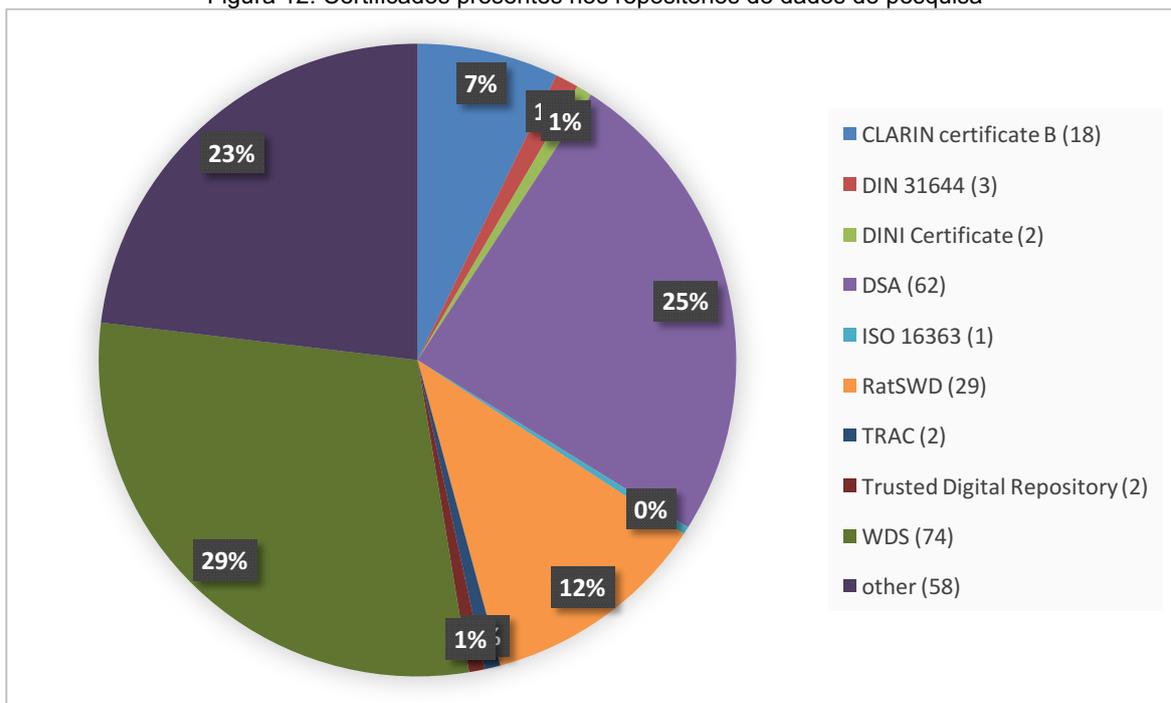


Fonte: Elaboração própria

Certificados

Cerca de 17% dos repositórios apresentavam algum tipo de certificado de sistema. Entre eles, destacaram-se o World Data System (WDS), que é amplamente utilizado por repositórios relacionados às ciências naturais, exatas e da terra; e o Data Seal of Approval (DSA), que é geralmente aplicado em repositórios cujo foco são as ciências sociais e humanas (Figura 12).

Figura 12: Certificados presentes nos repositórios de dados de pesquisa



Fonte: Elaboração própria

O WDS é uma iniciativa do International Council for Science (ICSU) e foi criado em 2008. Seu objetivo é promover a governança de longo prazo e o acesso universal e equitativo a dados científicos com garantia de qualidade, além de serviços de dados, produtos e informação.

O DSA é um selo desenvolvido pela Data Archiving and Network Services (DANS), um instituto criado, em 2009, pela Academia Real de Arte e Ciência dos Países Baixos (KNAW) com apoio da Organização para Pesquisa Científica dos Países Baixos (NWO). Seu objetivo é resguardar dados, garantindo sua qualidade, bem como assegurar o gerenciamento confiável dos dados para uso futuro sem a necessidade de implementar novos padrões ou onerar sua utilização.

O RatSWD é um conselho independente de diversos pesquisadores de universidades, institutos de pesquisa e produtores de dados da Alemanha. Foi criado em 2004 pelo Ministério Alemão de Educação e Pesquisa, e seu objetivo é melhorar a pesquisa em infraestrutura de dados e sua competitividade internacional.

Ao se analisarem os três certificados mais utilizados, é possível perceber que eles possuem características bem peculiares, apesar de algumas similaridades. Do ponto de vista de abrangência, o RatSWD certifica apenas repositórios que possuem participação

alemã em sua constituição. Já o DSA e o WDS certificam repositórios independentemente dos países em que foram criados. Outra peculiaridade pode ser identificada ao se analisar o processo de criação desses certificados. Enquanto o WDS foi criado a partir de um consórcio internacional, o DSA e o RatSWD possuem sua criação baseada, respectivamente, nos Países Baixos e na Alemanha. O DSA foi fomentado por entidades de governo e o RatSWD, apesar de possuir respaldo do Ministério Alemão para Educação e Pesquisa, também possui instituições privadas ligadas a sua criação.

Por fim, podem-se analisar os três certificados quanto às temáticas dos repositórios depositados. Enquanto 92% dos repositórios que utilizam o WDS armazenam dados relacionados às ciências naturais ou ciências exatas e da terra, o DSA certifica 94% de repositórios cujo foco é em ciências humanas, 45% em letras e linguística e 29% em ciências naturais. Já os repositórios certificados pelo RatSWD possuem seu foco principalmente em ciências humanas (72%), sociais aplicadas (59%), e saúde (21%).

Software

As informações acerca do software dos repositórios de dados de pesquisa cadastrados no re3data.org foram consideradas as menos abrangentes obtidas neste estudo. Apenas 12% dos registros apresentaram alguma indicação a respeito dos programas computacionais utilizados para gerenciar a base de dados de seus sistemas. Parte dos sistemas que não apresentaram informações relacionadas ao software (22% do total) indicou que não havia opções pertinentes sobre o tema no momento do registro e, portanto, responderam que utilizam “outro” software entre os listados.

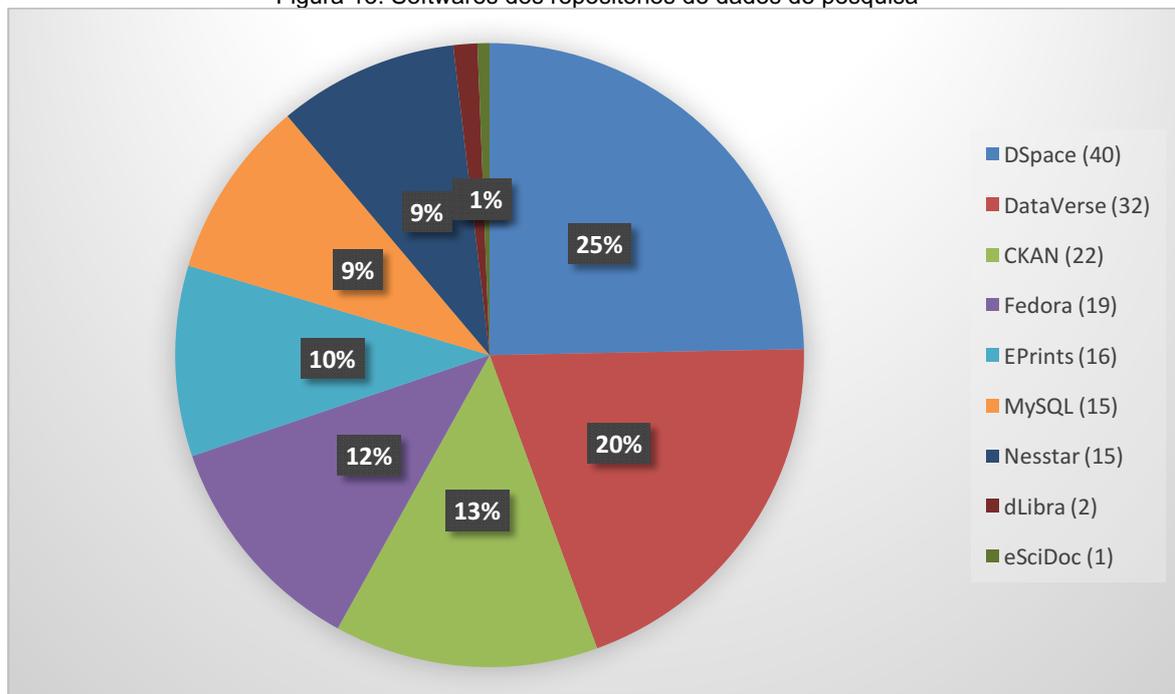
Entre os repositórios que indicaram o software utilizado, a maioria (25%) está relacionada ao uso do DSpace (Figura 13). O software DSpace permite o acesso a todos os tipos de conteúdo digital e possui sua utilização focada em academia e organizações comerciais e sem fins lucrativos. Ele é gratuito e de código aberto e pode ser facilmente adaptado para atender as especificidades do repositório. Ele é mantido pela DuraSpace, uma organização sem fins lucrativos.

O segundo software mais utilizado para criação de repositórios de dados de pesquisa, identificado pelo estudo, foi o DataVerse, que está presente em 20% dos sistemas que apresentaram informação acerca do software. O DataVerse é um software de código aberto voltado para preservação, citação, exploração e análise de dados de pesquisa. Foi idealizado através da parceria entre o Instituto para Ciência Social

Quantitativa e a Universidade de Harvard. Além do software, essa parceria permitiu a criação de um programa de suporte na utilização do sistema.

O software CKAN foi identificado como o terceiro mais utilizado entre os sistemas analisados, estando presente em 13% dos repositórios de dados de pesquisa. O CKAN é mantido pela Open Knowledge Foundation e possui o propósito de compartilhar, publicar, encontrar e utilizar dados. Sua utilização é focada em governos e iniciativa privada.

Figura 13: Softwares dos repositórios de dados de pesquisa



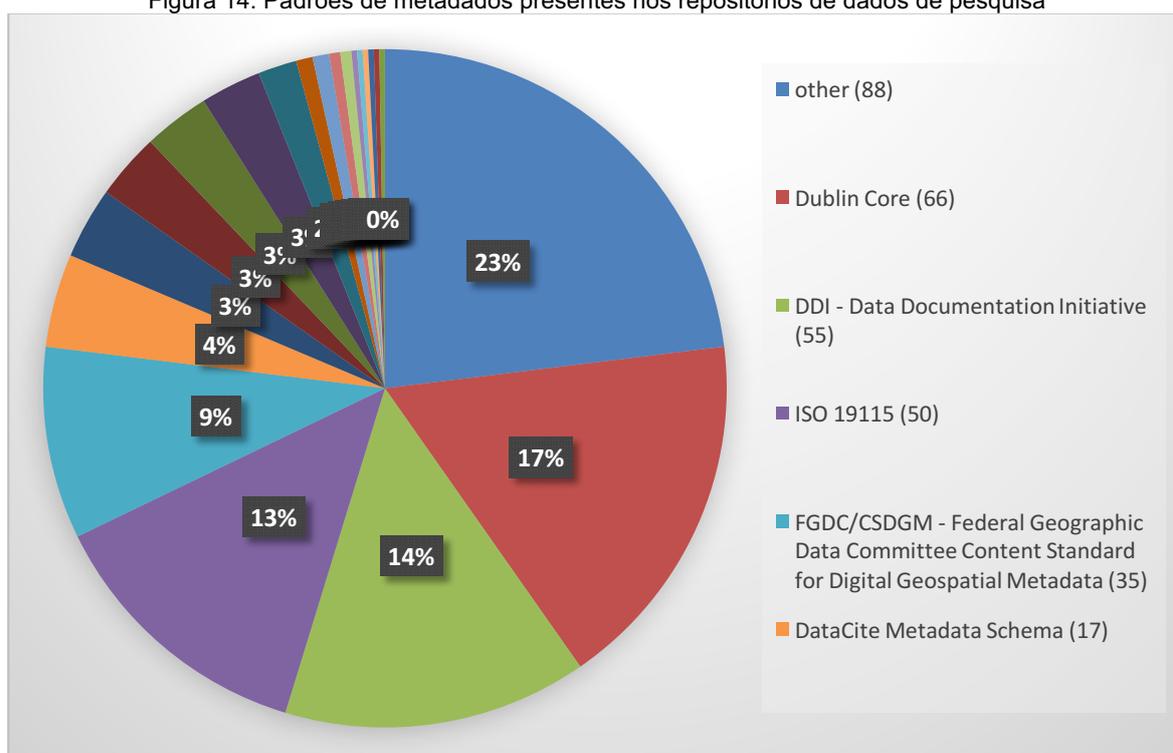
Fonte: Elaboração própria

De forma geral, o software realiza basicamente a mesma função de gestão de dados de pesquisa. Enquanto o DSpace se coloca claramente focado no ambiente acadêmico, o CKAN e o DataVerse não enfatizam esse uso de forma tão ativa. No que se refere à mantenedora do software, tanto o DSpace quanto o CKAN são mantidos por instituições sem fins lucrativos, enquanto o DataVerse possui uma grande instituição mantenedora, a Universidade de Harvard. Uma das grandes características do DSpace é a possibilidade de adaptação do sistema às necessidades do usuário e sua relação com uma grande comunidade de usuários e desenvolvedores.

Metadados

No que diz respeito à característica “metadados”, foi possível compreender, de forma não abrangente, os padrões utilizados pelos repositórios. Do total de repositórios disponíveis no diretório re3data, apenas 24% deles apresentam informações relativas aos metadados utilizados para a descrição dos seus recursos. Destes destacam-se o Dublin Core, DDI e a International Organization for Standardization (ISO) 19115. Destaca-se aqui que 23% dos repositórios que possuem categorização por metadados não apresentam a definição exata de qual padrão utilizam, sendo categorizados como outros (Figura 14).

Figura 14: Padrões de metadados presentes nos repositórios de dados de pesquisa



Fonte: Elaboração própria

O padrão Dublin Core está presente em 17% dos repositórios que apresentam informações sobre seus metadados. Segundo informações apresentadas no DCMI (2016), o padrão possui cinco grandes princípios: construção de forma aberta e sob consenso; participação e escopo internacional; neutralidade em seus propósitos e modelos de negócio; neutralidade na tecnologia e foco transdisciplinar. A organização e a operação da iniciativa são formalizadas por meio de uma série de regras que definem como os membros

da iniciativa atuarão no que diz respeito a suas responsabilidades e aos processos de decisão sobre o padrão.

Já o padrão DDI é focado na descrição de dados provenientes das ciências sociais, comportamentais e econômicas (KRAMER; LEAHEY, 2012) e está presente em 14% dos repositórios que possuem essa característica. Esse padrão é expresso em XML e contempla todo o ciclo de vida dos dados de pesquisa, desde a etapa de conceituação, passando pela análise e arquivamento dos dados. O projeto de criação do padrão DDI remonta ao ano de 1995, e a primeira versão foi apresentada à comunidade em idos de 2000. Desde então o padrão passou por diversos processos de redesenho, e a versão atual é a DDI Lifecycle 3.2, que possui características modulares e a capacidade de ser expandida (DATA DOCUMENTATION INITIATIVE, 2013).

Por fim, o padrão ISO 19115 é aplicado em 13% dos repositórios. O padrão foi atualizado pela última vez no ano de 2014 e seu foco é a descrição de informações geográficas. Segundo a ISO (2014), o padrão é aplicável à catalogação de todos os tipos de recursos, bem como serviços geográficos, conjunto de dados e características geográficas.

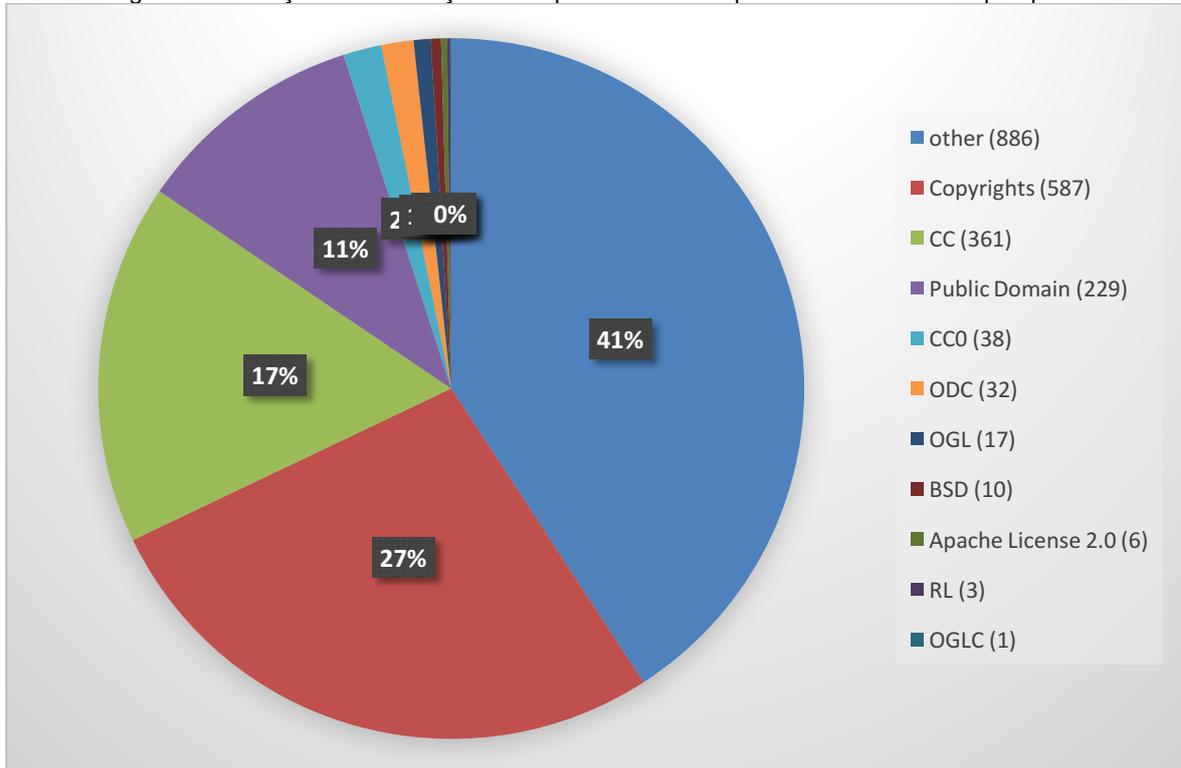
Ao se analisarem os três padrões de metadados, percebe-se que o Dublin Core possui características mais abrangentes em relação a sua utilização. O DDI também possui essa característica, mas o seu processo de criação foi focado em áreas específicas do conhecimento. Por fim, o padrão ISO 19115, embora possua a indicação de uso para qualquer tipo de documento, tem seu desenvolvimento focado em dados geográficos, o que indica que o padrão é o mais específico entre os três.

2.2.2.3. Direitos de propriedade intelectual nos repositórios de dados de pesquisa

A última categoria de análise avaliou o comportamento dos repositórios em relação aos princípios da ciência aberta declarados pelo Panton Principles (2011), no qual a abertura dos dados significa a disponibilização livre na Internet, com permissões para download, cópia, análise, reprocessamento e uso em softwares. As permissões devem ser explicitamente declaradas por meio de uma licença adequada (MOLLOY, 2011). De forma geral, os repositórios analisados não foram satisfatórios em relação à apresentação de licença para a ciência aberta, pois 41% deles não apresentaram informações sobre a licença utilizada e 27% declararam utilizar a licença copyright. Apenas um quarto dos

repositórios cumpriu a adequação proposta pelo Panton Principles, 17% utilizando licenças Creative Commons e 11% empregando licenças Public Domain (Figura 15).

Figura 15: Licenças de distribuição e uso presentes nos repositórios de dados de pesquisa



Fonte: Elaboração própria

Apesar de o tema sobre a comunicação dos dados de pesquisa ter sobressaído recentemente nas discussões sobre comunicação científica e ciência aberta, sua prática não é nova. As primeiras tecnologias e serviços de informação para essa finalidade datam da década de 1950, a partir da criação do World Data Centre (WDC). Para compreender o desenvolvimento da discussão, realizou-se um levantamento histórico das principais iniciativas e princípios orientadores para a comunicação dos dados de pesquisa.

2.2.3. Breve histórico das principais iniciativas relacionadas aos dados de pesquisa¹³

Para a narrativa histórica proposta, o WDC foi considerado a primeira iniciativa marcante consolidada com a finalidade de comunicação de dados de pesquisa. O WDC¹⁴ foi criado em 1957 para armazenar e dar acesso aos dados de um programa de pesquisa na área da geofísica. Inicialmente foram estabelecidos centros nos Estados Unidos, Europa, Japão e União Soviética. Desde então, segundo Ruttenberg e Rishbeth (1994), o sistema tem sido utilizado por projetos internacionais de diversas áreas do conhecimento, e já foram criados mais de 52 centros em 12 países.

Quinze anos mais tarde, em 1971, a United Nations Educational, Scientific and Cultural Organization (UNESCO) publicou relatório de estudo sobre a viabilidade de um sistema de informação científica mundial, denominado *United Nations International Scientific Information System* (UNISIST). O modelo UNISIST teve sua relevância destacada por Søndergaard et al. (2003) por causa de sua perspectiva sociotécnica das atividades de comunicação científica, que previa a interação entre produtores e usuários de informação científica e técnica por meio de diferentes canais de comunicação: informais, formais e tabulares. Os dois primeiros referiam-se às interações pessoais e publicações/literatura científica, respectivamente. Os canais tabulares, por sua vez, foram previstos no modelo UNISIST com a finalidade de comunicar dados de pesquisa por meio de unidades organizacionais denominadas de centros de dados. No contexto em que o modelo foi proposto, os dados seriam apresentados em forma de tabelas e, embora seja limitado para a atuação contemporânea da ciência, já previa uma alternativa ao fluxo linear do texto falado ou escrito das categorias anteriores (UNISIST, 1971; SØNDERGAARD et al., 2003). A relevância do modelo para as iniciativas pioneiras de comunicação dos dados de pesquisa reside na perspectiva visionária que ele apresenta de que há necessidade de considerar dados de pesquisa no fluxo da comunicação da ciência, conforme destacou Leite (2011).

Na década de 1980, iniciou-se uma discussão internacional a respeito da coleta e organização de dados sobre o *deoxyribonucleic acid* (DNA). Apesar da ampliação da discussão em âmbito internacional, não foi localizada em seus documentos nenhuma referência às duas iniciativas anteriores, o WDC e o UNISIST. No fim da década, o governo

¹³ Os textos das seções 2.2.3 e 2.2.4 foram publicados no seguinte artigo:

COSTA, M.; LEITE, F. *Princípios e recomendações basilares para a comunicação dos dados de pesquisa*. Em *Questão*. v. 23, n. 1, jan/abr. 2017. Disponível em: <http://www.seer.ufrgs.br/EmQuestao/article/view/65623>.

¹⁴ Página web do sistema de informação: <https://www.icsu-wds.org/services/data-portal>

norte-americano encomendou um planejamento do projeto para a execução do seu mapeamento. O PGH iniciou em 1990 e contemplou a coleta de dados sobre o gene humano, os métodos para o armazenamento dos dados e o desenvolvimento de ferramentas necessárias para analisá-los. A primeira fase do PGH foi desenvolvida com a participação de pesquisadores do Reino Unido, Alemanha, França, Japão e China, e passou a ser gerido por uma instituição internacional. O PGH recebeu investimento de 3 bilhões de dólares e foi dado por concluído em 2003, 2 anos antes do prazo previsto. Segundo Zatz (2000), o objetivo do PGH era “identificar todos os genes responsáveis por nossas características normais e patológicas”. A expectativa desse projeto era, a partir dos genes mapeados, desenvolver vacinas de DNA, que teriam alta eficiência na cura de doenças. Zatz, que em 2000 era coordenadora do Centro de Estudos do Genoma Humano, expectou que “as vacinas de DNA poderiam eliminar doenças como a tuberculose ou a Aids”.

Em meados da década de 1990, o U.S. Global Change Research Program (USGCRP) iniciou um debate público sobre o princípio do compartilhamento de dados de pesquisa. O programa norte-americano iniciou-se na década de 1990, com mandato presidencial, para desenvolver pesquisas sobre as mudanças climáticas (RAND, 1995). Cerca de 5 anos após a criação do programa, a organização já estava convencida da necessidade do compartilhamento de dados para o sucesso dos estudos.

Os Estados Unidos da América ainda deram outro passo importante para a comunicação dos dados de pesquisa no campo das ciências da saúde em 2000, quando o NIH, agência de fomento à pesquisa norte-americana, instituiu sua política de comunicação dos dados de pesquisa. O NIH determinou, mediante política institucional, que, a partir de outubro de 2003, todo o projeto de pesquisa financiado pela instituição deveria ser acompanhado de planejamento para o compartilhamento dos dados de pesquisa. Em linhas gerais, a política determina que “os dados devem estar disponíveis livremente o mais rápido possível, salvaguardando a privacidade dos participantes, e protegendo dados confidenciais e proprietários” (NATIONAL, 2003).

A iniciativa norte-americana parece ter impulsionado outros países. Em 2004, a OECD publicou o documento intitulado *Declaration on Access to Research Data from Public Funding*, que foi assinado pelo governo de 34 países. Na declaração, reafirma-se a importância da comunicação dos dados de pesquisa para o desenvolvimento da ciência e otimização dos investimentos públicos em pesquisa científica (OECD, 2004). No documento, são declarados dez princípios a serem seguidos pelos países signatários (Quadro 3). Neste documento, observa-se o mesmo argumento incluído na política do NIH:

tornar públicos e acessíveis os dados resultantes de pesquisas financiadas com recursos públicos, para ampliar o desenvolvimento científico e conferir transparência aos investimentos na pesquisa.

Quadro 3: Princípios da declaração da OECD (2004).

Princípio	Definição
Abertura	Equilibrar os interesses do acesso aberto aos dados a fim de aumentar a qualidade e a eficiência da pesquisa e inovação, com a necessidade de restrição de acesso, em alguns casos, para proteger os interesses sociais, científicos e econômicos.
Transparências	Tornar disponíveis e acessíveis em âmbito internacional a informação sobre as organizações de produtores de dados, a documentação que eles produzem e as especificações das condições associadas ao uso desses dados.
Conformidade legal	Prestar a devida atenção à concepção de regimes de acesso para dados de pesquisa, observando os requisitos legais nacionais em matéria de segurança nacional, privacidade e segredos comerciais.
Responsabilidade formal	Promover regras institucionais explícitas e formais sobre as responsabilidades das várias partes envolvidas em atividades relacionadas a informações relativas à autoria, créditos de produção, propriedade, restrições de uso, disposições financeiras, regras éticas, termos de licenciamento e responsabilidade.
Profissionalismo	Elaborar regras institucionais para a gestão de dados de pesquisa digitais com base nas normas e valores profissionais consagrados nos códigos de conduta das comunidades científicas envolvidas.
Proteção da propriedade intelectual	Descrever maneiras de obter acesso aberto no âmbito dos diferentes regimes jurídicos de direito autoral ou outro direito de propriedade intelectual aplicável às bases de dados e aos segredos comerciais.
Interoperabilidade	Prestar a devida atenção aos requisitos de padrões internacionais relevantes para o uso de várias formas, em cooperação com outras organizações internacionais.
Qualidade e segurança	Descrever boas práticas para métodos, técnicas e instrumentos utilizados na coleta, difusão e arquivamento acessível de dados para permitir o controle de qualidade por meio de análise pelos pares e outras formas de preservar a autenticidade, originalidade, integridade, segurança e de estabelecer as responsabilidades.
Eficiência	Promover maior eficácia de custo dentro do sistema mundial da ciência, descrevendo boas práticas de gestão de dados e serviços de apoio especializados.
Prestação de contas	Avaliar o desempenho dos regimes de acesso de dados para maximizar o apoio ao acesso aberto entre a comunidade científica e a sociedade em geral.

Fonte: (OECD, 2004) (traduzido e com adaptações).

Três anos depois, em 2007, a OECD publicou o segundo documento, denominado *OECD Principles and Guidelines for Access to Research Data from Public Funding*, cujo objetivo é traçar amplo quadro de recomendações para as políticas científicas dos países-membros, para que os dados resultantes de pesquisas financiadas com recursos públicos estejam amplamente disponíveis e não desrespeitem as políticas específicas de cada país. Quanto às singularidades regionais, o documento também declara que não pretende aumentar as obrigações dos países nem adicionar novos custos aos seus sistemas de

produção de conhecimento científico (OECD, 2007). No novo documento, são listados os dez princípios propostos em 2004 e os modos de executá-los. No entanto, o princípio Qualidade e Segurança foi dividido em dois, um deles trata apenas da qualidade e o outro da segurança. Ademais, foram acrescentados mais dois itens: flexibilidade e sustentabilidade (Quadro 4).

Quadro 4: Novos princípios propostos pela OECD (2007).

Princípio	Definição
Qualidade	Este princípio é indicado como o elemento que conferirá possibilidade de novos usos para os dados. Para tanto, o documento recomenda a adequação a padrões internacionais de qualidade. No entanto, destaca que esses padrões de qualidade não devem ser universais e devem respeitar as especificidades de cada área do conhecimento.
Segurança	Neste princípio, mostra-se a necessidade de suporte e instrumentos que garantam a integridade, a completude dos dados e a ausência de erros.
Flexibilidade	O princípio levanta a importância de considerar os avanços tecnológicos que podem afetar o armazenamento e o acesso dos dados. Também aponta a conformidade com os sistemas jurídicos e culturais de cada país-membro.
Sustentabilidade	O último princípio aborda aspectos sobre o financiamento do projeto de pesquisa que deve considerar a manutenção em longo prazo dos dados resultantes da pesquisa.

Fonte: Elaboração própria.

A ampliação da noção da abertura dos dados em relação à filosofia aberta deu-se em 2009, quando foi criado um grupo de trabalho visando à comunicação de dados no contexto da ciência. O grupo, intitulado Work Grouping Open Data in Science (WGOADS), foi articulado dentro da Open Knowledge Foundation (OKF), que tem por objetivo facilitar o acesso e o uso de dados em vários contextos, inclusive na ciência. Seguindo os princípios da fundação, o grupo declarou ter por missão orientar e criar ferramentas que possibilitem o compartilhamento e o reuso dos dados de pesquisa. De acordo com Molloy (2011), coordenadora do WGOADS, apesar dos problemas culturais e da inércia ante a ciência aberta, renomadas agências de fomento à pesquisa têm requisitado um plano para o compartilhamento dos dados como condição para o financiamento de um estudo. Nesse sentido, a autora destaca a iniciativa do NIH e da Wellcome Trust como fortes exemplos dessa tendência.

A fim de oferecer suporte aos pesquisadores e possíveis usuários dos dados compartilhados, o grupo desenvolveu a ferramenta denominada *Is It Open Data?*¹⁵. Segundo sua própria apresentação, a ferramenta visa disponibilizar informações sobre a abertura dos dados, em especial os dados de pesquisa. O sistema considera como dados de pesquisa todos os dados primários relacionados a uma publicação científica revisada

¹⁵ Fonte: <http://www.isitopendata.org/>

pelos pares (*paper*). Além dessa ferramenta, o WGODS sistematizou um conjunto de condições que define a abertura do acesso e do uso dos dados (Quadro 5).

Quadro 5: Condições de abertura dos dados propostos pelo WGODS.

Princípios	
1.	É livre para uso e reuso.
2.	É distribuído sem restrições.
3.	Utiliza a licença que determina que dados derivados compartilhem a mesma licença de abertura.
4.	É disponível em todo o seu conjunto.
5.	Quando possuir custo de reprodução, este deve ser razoável.
6.	Preferencialmente deve ser apto para o acesso por meio da Internet e sem custos.
7.	É disponível em formato conveniente e modificável.

Fonte: Elaboração própria (traduzido e adaptado de MOLLOY, 2011).

As condições determinadas pelo grupo foram baseadas na definição de dados abertos da OKF e do Protocol for Implementing Open Access Data¹⁶. Molloy (2011) justifica que a definição de abertura para os dados foi necessária dentro da perspectiva da ciência aberta, porque se considerou que a noção aplicada às publicações científicas pelo Movimento de Acesso Aberto não produzia liberdade de acesso ao conhecimento. A autora defende que a ideia de “acesso aberto” (*free access* ou *open access*) está relacionada à perspectiva de acesso gratuito, e não aberto, como almejam o grupo e seus parceiros. Molloy (2011) defende seu argumento mostrando dados dos periódicos científicos de acesso aberto cadastrados no DOAJ. Segundo ela, apenas 22% dos periódicos lá cadastrados utilizam as licenças Creative Commons, e desse conjunto alguns periódicos ainda contêm restrições quanto ao uso comercial e não demandam que obras derivadas compartilhem da mesma licença. Com isso, apenas uma reduzida parte dos periódicos classificados como de acesso aberto promovem de fato o conhecimento aberto.

Diante da problemática traduzida pelo grupo, Molly (2011) observa que se fez necessária a criação de um conjunto de princípios para a comunicação dos dados de pesquisa. O novo conjunto de princípios, intitulado *Panton Principles: Principles for Open Data in Science*, foi elaborado por quatro pesquisadores, três do Reino Unido e um dos Estados Unidos, e é assinado por mais de 265 pesquisadores¹⁷. O documento parte da ideia geral de que a ciência é baseada na reutilização e discussão do conhecimento científico já publicado, e que esse processo é mais efetivo se os dados das publicações estiverem abertos, ressaltando assim os princípios da ciência moderna. Para a autora, a abertura dos dados significa a disponibilização livre na Internet com permissões para

¹⁶ Fonte: <http://sciencecommons.org/projects/publishing/open-access-data-protocol/>

¹⁷ Dados de dezembro de 2015.

download, cópia, análise, reprocessamento e uso em software. Concluindo que os dados precisam ser disponibilizados em um local de domínio público. Após a definição de contexto e de conceitos, são propostos quatro princípios para a abertura dos dados de pesquisa (Quadro 6).

Quadro 6: Princípios declarados no Panton Principles.

Princípios	
1.	Os dados devem ser acompanhados de uma declaração clara e explícita sobre suas permissões de uso. Essa declaração precisa ser clara, irrevogável e baseada em norma jurídica apropriada.
2.	As permissões de uso devem ser declaradas por meio de licença adequada.
3.	A utilização da definição de dados proposta pela Open Knowledge não limita o uso comercial dos dados e outras restrições de uso.
4.	A permissão do uso dos dados, principalmente daqueles financiados com recurso público, deve ser feita com licenças de domínio público, tais como a PDDL ou CC0 (apresentadas a seguir).

Fonte: Panton Principles (traduzido e adaptado).

Nos princípios enunciados sobressai a preocupação com o licenciamento dos dados de pesquisa para o contexto da ciência aberta. De acordo com os Panton Principles, apesar da proximidade com a filosofia e os princípios do software livre, para o contexto de dados de pesquisa é desencorajado o uso das mesmas licenças utilizadas pelos softwares, tais como a Free Documentation License (GFDL), a GNU General Public License (GPL) e a Berkeley Software Distribution (BSD). Para a adequação da licença e da definição de dados proposta pela Open Knowledge, no contexto da ciência aberta o documento traz uma lista de seis licenças adequadas à comunicação de dados de pesquisa (Quadro 7).

Quadro 7: Licenças recomendadas pelo Panton Principles.

Licenças	Domínio	BY*	SA**
Creative Commons CCZero (CC0)	Conteúdo e dados	N	N
Open Data Commons Public Domain Dedication and License (PDDL)	Dados	N	N
Creative Commons Attribution 4.0 (CC-BY-4.0)	Conteúdo e dados	S	N
Open Data Commons Attribution License (ODC-BY)	Dados	S	N
Creative Commons Attribution Share-Alike 4.0 (CC-BY-SA-4.0)	Conteúdo e dados	S	S
Open Data Commons Open Database License (ODbL)	Dados	S	S

Fonte: <http://opendefinition.org/licenses/#Data> (traduzido e adaptado).

*BY – a licença requer atribuição de autoria.

**SA – produtos derivados dos dados cobertos pela licença deverão compartilhar da mesma licença.

Sobre o conjunto de licenças, ressaltam-se três características. A primeira trata do amplo reuso das licenças, uma vez que elas não são especificadas por organização ou jurisdição. A segunda é acerca das aproximações da licença com outras da mesma natureza, tais como as utilizadas pelos softwares livres. A terceira aborda a atualização e

as recomendações direcionadas às licenças por projetos que objetivam compartilhar seus dados.

No mesmo ano da publicação do Panton Principles, organizou-se um novo grupo de trabalho chamado Publishing Open Data. De acordo com Hrynaszkiewicz e Cockerill (2012), o grupo foi articulado pelos editores da BioMed Central para abrir um canal de discussão com outros editores acadêmicos sobre o compartilhamento de dados das pesquisas publicadas por seus periódicos. Como resultado dessa articulação, foi publicada, em 2011, uma política intitulada Join Data Archiving Policy. No documento, são descritos os requerimentos para a publicação de dados em artigos acadêmicos em várias áreas do conhecimento. Adicionalmente, são apontados os elementos que uma política editorial deve abarcar para requerer o compartilhamento dos dados dos estudos que publicam. A política também destaca o Dryad Digital Repository¹⁸, que armazena dados de pesquisas na área de ciências da saúde.

Ainda em 2011, os Research Councils United Kingdom (RCUK)¹⁹ publicaram um novo conjunto de princípios para a comunicação de dados de pesquisa, o *RCUK Common Principles on Data Policy* (Quadro 8). De acordo com Corti et al. (2014), tal política representou grande avanço na promoção do compartilhamento de dados de pesquisa, porque o RCUK é responsável pelo financiamento de 55% das pesquisas da região e influenciou o comportamento de outras instituições fora de seu domínio.

Quadro 8: Princípios para a comunicação dos dados de pesquisa da RCUK.

Princípio	
1.	Dados de pesquisa resultantes de financiamento público são um bem público, produzido para o interesse público. Assim, devem ser disponibilizados abertamente, com o menor número de restrições possível, de forma oportuna e responsável.
2.	Políticas e planos de gerenciamento de dados institucionais e projetos específicos devem estar em conformidade com as normas e melhores práticas da sua comunidade científica. Dados com valor de longo prazo devem ser preservados e permanecerão acessíveis e utilizáveis para as pesquisas futuras.
3.	Para permitir que os dados da pesquisa sejam detectáveis e efetivamente reutilizados por outras pesquisas, seus metadados devem ser registrados de maneira adequada e devem estar abertamente disponíveis para permitir que outros pesquisadores compreendam o potencial dos dados de pesquisa e as possibilidades de reuso. Os resultados publicados devem sempre incluir informações sobre como acessar os dados de apoio.

(continua)

¹⁸ Fonte: <http://datadryad.org/>

¹⁹ O RCUK é uma organização que reúne sete conselhos de pesquisa do Reino Unido e, segundo seus dados, investe 3 bilhões de libras em pesquisa por ano. Fonte: <http://www.rcuk.ac.uk/about/aboutrcuk/rcuk-ambition/>

Princípio	
4.	RCUK reconhece que há restrições legais, éticas e comerciais sobre o acesso aos dados de pesquisa. Para garantir que o processo de pesquisa não seja danificado pelo lançamento inadequado de dados, políticas e práticas de organização de pesquisa devem garantir que essas restrições serão consideradas em todas as fases do processo de pesquisa.
5.	Para assegurar que as equipes de pesquisa obtenham o reconhecimento apropriado para o esforço envolvido na recolha e análise de dados, poderá ser dado o direito a um período limitado de utilização privilegiada dos dados que foram coletados para capacitá-los a publicar os resultados de sua pesquisa. A duração desse período variará de acordo com a disciplina de investigação e, se for caso, será analisado nas políticas publicadas pelos Conselhos de Pesquisa.
6.	A fim de reconhecer as contribuições intelectuais dos pesquisadores que geram, preservam e compartilham conjuntos de dados de pesquisa, todos os usuários de dados da pesquisa devem reconhecer as fontes de seus dados e respeitar os termos e as condições sob as quais são acessados.
7.	É conveniente utilizar recursos públicos para apoiar a gestão e partilha de dados da pesquisa financiada publicamente. Para maximizar o benefício de pesquisa, que pode ser adquirida a partir de orçamentos limitados, os mecanismos para essas atividades devem ser eficientes e utilizar recursos públicos rentáveis.

Fonte: RCUK, 2011 (traduzido).

Segundo Corti et al. (2014), até 2014, onze instituições de pesquisas já haviam instituído sua política de comunicação de dados de pesquisa, baseadas nos princípios propostos pelo RCUK. As políticas determinam que um plano para o compartilhamento dos dados resultantes da pesquisa deve ser entregue junto com a proposta da pesquisa. Assim, o planejamento para a comunicação dos dados da pesquisa faz parte do próprio planejamento da pesquisa. Para viabilizar o planejamento, é oferecido aos pesquisadores um conjunto de serviços de suporte e infraestrutura. Corti et al. (2014) destacaram cinco deles:

- UK Data Service do Economic and Social Research Council (ESRC)²⁰.
- Rede de dados do Natural Environment Research Council (NERC)²¹.
- MRC Data Support Service do Medical Research Council (MRC)²².
- Atlas Petabyte Storage do Science and Technology Facilities Council (STFC)²³.
- Archeology Data Service do Arts and Humanities Research Council (AHRC)²⁴.

Basicamente, todos os serviços disponibilizam orientações para a gestão e o compartilhamento dos dados e oferecem suporte para a curadoria, preservação e compartilhamento dos dados de pesquisa para as áreas de conhecimento em que trabalham.

No contexto europeu, destaca-se o programa Horizon 2020 para promoção da comunicação de dados de pesquisa. O programa tem por objetivo otimizar os investimentos

²⁰ <https://www.ukdataservice.ac.uk/>

²¹ <http://www.nerc.ac.uk/research/sites/data/>

²² <http://www.mrc.ac.uk/research/research-policy-ethics/data-sharing/data-support-service/>

²³ <http://www.scd.stfc.ac.uk/e-Science/services/atlas-petabyte-storage/22459.aspx>

²⁴ <http://archaeologydataservice.ac.uk/>

da União Europeia em pesquisa e inovação. Para tanto, busca promover mais acesso e com mais rapidez aos resultados de pesquisa que financia. Os resultados são expostos em publicação científica e dados de pesquisa. O conjunto das publicações científicas abrange os documentos revisados pelos pares, tais como artigos de periódicos, e os não revisados, como os livros e relatórios de pesquisa. Já os dados de pesquisa

[...] referem-se às informações, em particular fatos ou números, coletados para ser examinados e considerados e como base para discussão ou cálculo. No contexto da pesquisa, exemplos de dados incluem estatísticas, resultados de experimentos, medidas, observações resultantes do campo de trabalho, resultados de questionários, gravações de entrevistas e imagens. O foco é os dados de pesquisa que estão disponíveis em formato digital (GUIDELINES, 2013).

O acesso aberto aos dados de pesquisa é definido como o direito de acesso e reuso dos dados sob os termos e condições do *Grant Agreement*²⁵. As condições estabelecidas pelo documento determinam que os beneficiários de recursos da União Europeia devem depositar seus dados de pesquisa em repositórios de dados apropriados, permitindo o seu acesso, mineração, exploração, reprodução e compartilhamento, sem custo para os usuários. No documento, declara-se a expectativa de que os dados de pesquisa sirvam como um objeto de validação dos resultados de pesquisa que foram ou serão publicados. Como condição para o depósito também se determina que o conjunto de dados seja estruturado de forma padronizada e os metadados incluam os seguintes elementos:

- a) os termos “European Union (EU)” e “Horizon 2020”;
- b) o nome do projeto financiado e seu número de controle;
- c) a data da publicação, o período de embargo, quando aplicável, e o identificador persistente (H2020, 2015).

Para o cumprimento dos requisitos, é necessário elaborar um plano de gestão dos dados. De acordo com o documento (GUIDELINES, 2013, p. 10),

[...] plano de gestão é um documento que esboça como o dado de pesquisa coletado ou gerado será manipulado durante um projeto de pesquisa, e depois dele completo, descrevendo quais dados serão coletados/gerados, seguindo qual metodologia e padrões, se e como esse dado será compartilhado e/ou será aberto para o acesso, e como será curado e preservado.

Apesar das exigências descritas, são consideradas possibilidades de exceção, ou seja, pesquisas que não terão seus dados disponibilizados em acesso aberto. As

²⁵ Os requisitos para a comunicação de dados no documento é apresentado no item 29.3.

pesquisas que se enquadram nessa categoria são aquelas passíveis de patente ou, por algum motivo, de proteção dos dados.

O início do ano de 2016 foi marcado pela ampla discussão que o tema do compartilhamento de dados de pesquisa recebeu em âmbito internacional. A visibilidade para o assunto decorreu do acordo de cooperação internacional denominado *Statement on Data Sharing in Public Health Emergencies*, que explicitou a necessidade do compartilhamento de dados de pesquisa para a superação de epidemias como o vírus da zika e o ebola. O documento foi assinado por 33 instituições, entre organismos internacionais, ONGs, institutos de pesquisas e grandes editores científicos (Quadro 9).

Quadro 9. Instituições aderentes ao acordo de cooperação internacional.

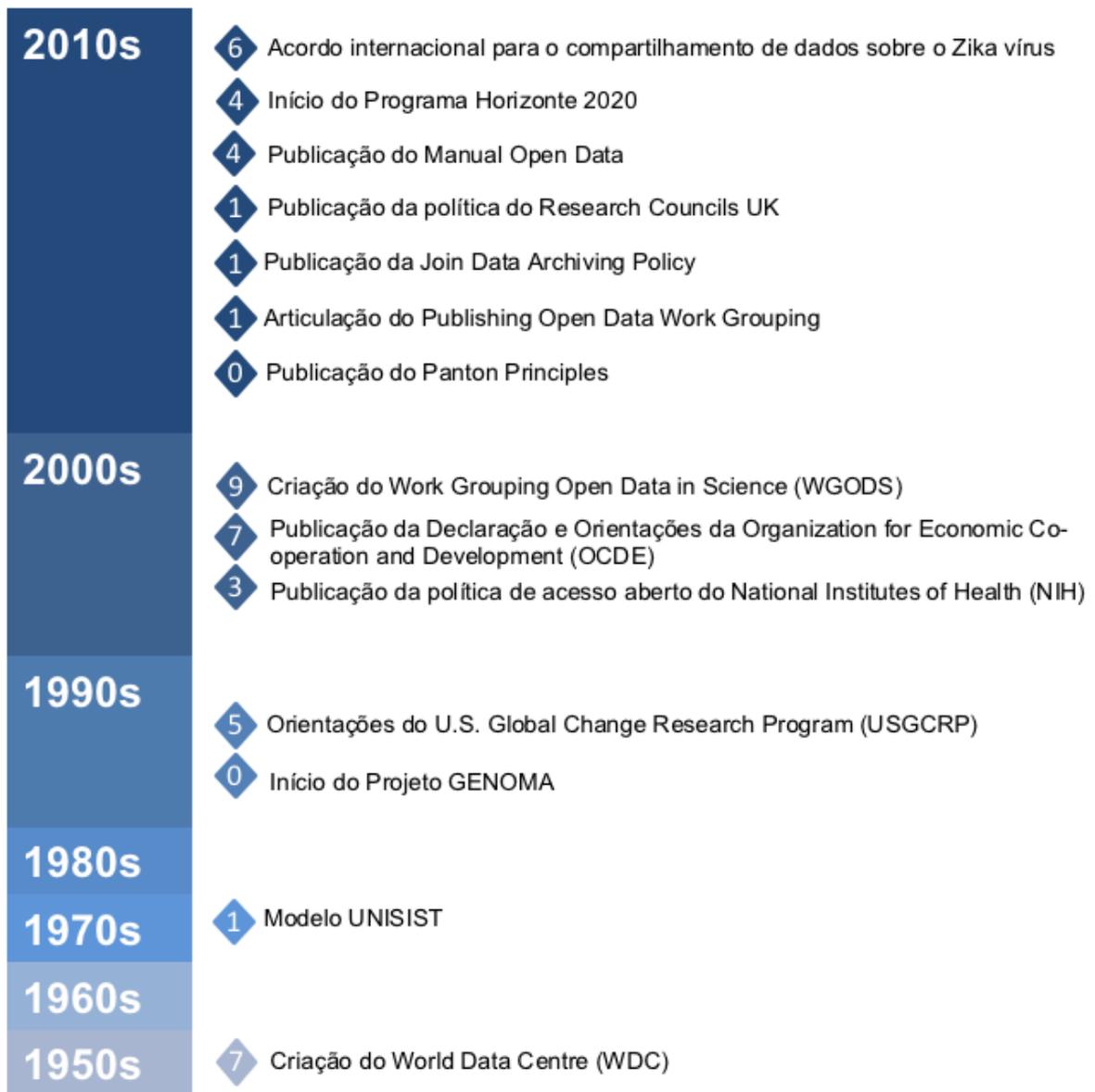
1.	Academy of Medical Sciences, UK
2.	Bill and Melinda Gates Foundation
3.	Biotechnology and Biological Sciences Research Council (BBSRC)
4.	The BMJ
5.	Bulletin of the World Health Organization
6.	Canadian Institutes of Health Research
7.	The Centers for Disease Control and Prevention
8.	Chinese Academy of Sciences
9.	Chinese Centre for Disease Control and Prevention
10.	The Department of Biotechnology, Government of India
11.	The Department for International Development (DFID)
12.	Deutsche Forschungsgemeinschaft (DFG)
13.	eLife
14.	The Economic and Social Research Council (ESRC)
15.	F1000
16.	Fondation Mérieux
17.	Fundação Oswaldo Cruz (Fiocruz)
18.	The Institut Pasteur
19.	Japan Agency for Medical Research and Development (AMED)
20.	The JAMA Network
21.	The Lancet
22.	Médecins Sans Frontières/Doctors Without Borders (MSF)
23.	National Academy of Medicine
24.	National Institutes of Health, USA
25.	National Science Foundation, USA
26.	The New England Journal of Medicine (NEJM)
27.	PLOS
28.	Science Journals
29.	South African Medical Research Council
30.	Springer Nature
31.	UK Medical Research Council
32.	Wellcome Trust
33.	ZonMw - The Netherlands Organisation for Health Research and Development

O objetivo declarado no acordo é garantir que qualquer informação com algum valor para o combate ao vírus da zika seja disponibilizada de modo rápido e gratuito. Além das doenças causadas pelo vírus da zika, o acordo realça que a prática permitirá responder com mais agilidade às futuras emergências da saúde pública. No acordo, são destacados dois compromissos assumidos pelos signatários, um para o grupo dos editores científicos

e outro para as instituições fomentadoras de pesquisa na área da ciência da saúde. O compromisso assumido pelos editores científicos é a disponibilização das publicações científicas e seus dados de pesquisa de forma rápida e aberta. Complementarmente, acordou-se com os editores que as pesquisas que tiverem seus dados ou *pre-prints* disponibilizados antes da publicação do artigo não terão sua submissão comprometida. Já as agências financiadoras de pesquisa se comprometeram a promover pesquisas relevantes para o tema das doenças emergentes e definir mecanismos que garantam a comunicação dos seus dados o mais rapidamente possível. O documento finaliza convidando outras instituições a aderirem ao acordo e disponibiliza, na própria página da web, um formulário para os interessados em se tornarem signatários.

Todos os eventos que foram marcados nesta seção estão ilustrados na Figura 16, no formato de linha do tempo.

Figura 16: Linha do tempo da promoção da abertura dos dados de pesquisa.



Fonte: Elaboração própria.

A sistematização da linha do tempo permitiu, dentre outros resultados, a identificação de cinco documentos que apresentaram princípios e recomendações para a comunicação dos dados de pesquisa. Tais documentos serão discutidos no próximo item desse capítulo.

2.2.4. Princípios e recomendações para a comunicação dos dados de pesquisa

Nesta seção, será apresentada uma síntese dos cinco documentos, com vistas à proposição de um conjunto de princípios acordados entre diversos países e instituições que já se dedicaram à discussão sobre o tema.

Em âmbito regional, a Europa predomina sobre as demais em quantidade de iniciativas e na adesão às declarações de princípios para a comunicação dos dados de pesquisa. Já os Estados Unidos não se destacam em relação à quantidade de declarações internacionais, mas notam-se a quantidade e a qualidade de iniciativas desenvolvidas naquele país, em especial na área de ciências da saúde. Na América Latina, observou-se a participação de apenas dois países, o Brasil e o México. A inclusão brasileira deu-se pelo envolvimento de uma pesquisadora na articulação WGODS. Nenhuma outra iniciativa brasileira foi identificada a partir da formalização proposta nesse recorte. A articulação do grupo WGODS também contou com a participação de pesquisadores de alguns países da Ásia e da Nova Zelândia. Já a adesão mexicana esteve relacionada à aderência da declaração da OECD, que, além de abranger países da América do Norte e Europa, também incluíram a África do Sul, a China, o Japão e a Coreia do Sul (Figura 17).

Figura 17: Mapa das declarações e princípios para a comunicação dos dados de pesquisa



Fonte: <https://goo.gl/qEshVS>

Em termos gerais, foi possível identificar 14 princípios declarados nos documentos como valores relevantes para o contexto da comunicação dos dados de pesquisa. Todos os quadros-resumos das declarações discutidas na seção anterior foram sistematicamente organizados no Quadro 10, a fim de comparar as propostas dos cinco documentos.

Quadro 10: Princípios e recomendações discutidos nas declarações analisadas.

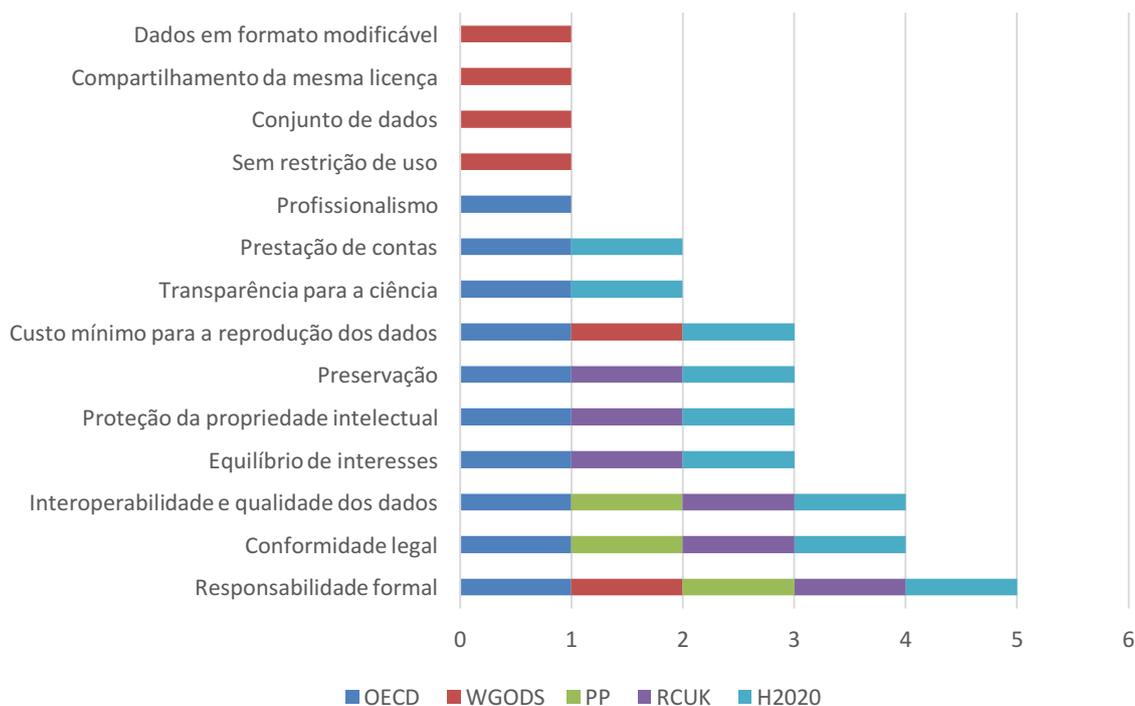
Princípios e recomendações	OECD	WGODS	PP²⁶	RCUK	H2020	Total
Responsabilidade formal	1	1	1	1	1	5
Conformidade legal	1	0	1	1	1	4
Interoperabilidade e qualidade dos dados	1	0	1	1	1	4
Equilíbrio de interesses	1	0	0	1	1	3
Proteção da propriedade intelectual	1	0	0	1	1	3
Preservação	1	0	0	1	1	3
Custo mínimo para a reprodução dos dados	1	1	0	0	1	3
Transparência para a ciência	1	0	0	0	1	2
Prestação de contas	1	0	0	0	1	2
Profissionalismo	1	0	0	0	0	1
Sem restrição de uso	0	1	0	0	0	1
Conjunto de dados	0	1	0	0	0	1
Compartilhamento da mesma licença	0	1	0	0	0	1
Dados em formato modificável	0	1	0	0	0	1
Total	10	6	3	6	9	x

Fonte: Elaboração própria.

Após a comparação, observou-se que apenas um dos 14 princípios foi declaradamente comum aos cinco documentos (Figura 18). Outros dois foram aventados em quase todos os documentos, exceto o WGods. Em relação à maioria dos princípios, notou-se a predominância dos seguintes documentos: declarações da OECD, política do RCUK e política do H2020. Os itens tratados no Panton Principles obtiveram menor ocorrência dentro do conjunto de princípios e recomendações observados, mas quando presentes acompanharam a tendência do grupo descrito anteriormente. O documento publicado pelo WGods destacou-se por incluir princípios singulares aos outros documentos.

²⁶ A sigla PP refere-se ao documento Panton Principles e foi utilizada apenas no Quadro 10 deste trabalho.

Figura 18: Comparação dos princípios e recomendações discutidos nas declarações analisadas.



Fonte: Elaboração própria.

O princípio comum a todos os documentos trata da formalização da responsabilidade em tornar o dado de pesquisa público. Com isso, os documentos buscam ser um instrumento para a promoção da comunicação dos dados de pesquisa. No entanto, a maneira como a abertura é tratada pelos documentos varia significativamente. Na declaração da OECD, é indicado o uso de regras institucionais que determinem as possíveis restrições de uso e os termos de licenciamento, entre outros aspectos. Já no documento do WGODS, é definido que a comunicação de dados de pesquisa dentro do contexto da ciência aberta não deve estipular restrições de uso e deve se utilizar de licenças internacionais adequadas para essa finalidade. De modo geral, essa variação marca a discrepância entre os princípios e recomendações contidos nos documentos analisados.

A declaração da OECD, a política do RCUK e o programa H2020 revelam tendências para mediação de interesses diversos relacionados aos resultados das pesquisas, considerando até mesmo atender restrições de acesso e reuso com vistas a interesses particulares, sejam eles comerciais ou de direito de propriedade intelectual. Já o documento do WGODS preza pela completa abertura do acesso e da reutilização dos dados de pesquisa, sem nenhuma restrição. Fundamentalmente os princípios que

divergem são os relacionados ao “equilíbrio de interesses” e à “proteção da propriedade intelectual”, que se opõem à “promoção do acesso e do uso sem restrições”.

Apesar da diferença apontada, foi possível chegar a um conjunto de princípios e recomendações que compõem o quadro das discussões sobre a comunicação dos dados de pesquisa em nível internacional (Quadro 11). O conjunto representará a base para as investigações propostas por esta pesquisa, servindo de categorias preliminares para a análise dos dados.

Quadro 11: Conjunto dos princípios e recomendações para a comunicação dos dados de pesquisa.

Princípios e recomendações		
1	Responsabilidade formal	Formalização do compromisso de tornar o dado de pesquisa público.
2	Conformidade legal	Destaca que a obrigatoriedade da comunicação de dados de pesquisa deve ser flexível aos sistemas jurídicos do país ou região.
3	Interoperabilidade e qualidade dos dados	Evidencia a necessidade de padrões internacionais para descrição, armazenamento e recuperação dos dados.
4	Equilíbrio de interesses	Considera exceções na obrigatoriedade do compartilhamento dos dados de pesquisa em função de interesses dos pesquisadores e financiadores.
5	Proteção da propriedade intelectual	Considera a possibilidade de restrição à comunicação dos dados de pesquisa em função dos direitos de propriedade intelectual e de segredos comerciais.
6	Custo mínimo para a reprodução dos dados	Determina que o acesso e o uso do dado de pesquisa gerem o mínimo de custo possível para o usuário.
7	Transparência para a ciência	Destaca a importância do acesso aos dados de pesquisa para tornar os resultados apresentados incorruptíveis e refutáveis.
8	Prestação de contas	Disponibilização dos dados de pesquisa como forma de comprovar a realização da pesquisa e o cumprimento de seus objetivos.
9	Sem restrição de uso	Recomenda que não seja atribuída nenhuma restrição de uso, nem mesmo de uso comercial.
10	Conjunto de dados	Requisição para todo o conjunto de dados produzidos e/ou utilizados pela pesquisa para a comprovação dos seus resultados.
11	Preservação	Necessidade de evidenciar formas e financiamentos para a preservação dos dados em longo prazo.
12	Profissionalismo	Uso de valores e técnicas profissionais para tratamento, armazenamento e disponibilização dos dados de pesquisa.
13	Compartilhamento da mesma licença	Implica a necessidade de que obras derivadas, ou que utilizem os dados de pesquisa, compartilhem da mesma licença, garantindo a abertura dos dados.
14	Dados em formato modificável	Indica que, sempre que possível, os dados de pesquisa devem ser compartilhados em formato modificável.

Fonte: Elaboração própria.

As propostas relacionadas à comunicação dos dados de pesquisa trazem como motivação central a promoção do avanço da ciência de maneira mais rápida e eficaz. Além dos argumentos em torno do amplo acesso público aos resultados de pesquisas financiadas com recursos públicos, identifica-se a expectativa da sua contribuição na resolução de problemas sociais e na democratização do acesso e da produção do conhecimento científico.

Chan e Costa (2005) acentuam a importância do acesso à ciência e à tecnologia para diminuir os problemas dos países em desenvolvimento. De acordo com os autores, tais países estão à margem da produção de ciência e tecnologia, em razão do baixo investimento em pesquisa nesses locais e das dificuldades de acesso à literatura científica internacional, em decorrência dos altos custos das assinaturas dos principais periódicos. Os autores relacionaram dados de 2004 publicados pela Organização Mundial da Saúde (OMS), que demonstravam que 56% dos países com Produto Interno Bruto (PIB) inferior a mil dólares não fizeram assinatura dos principais periódicos científicos, entre 1999 e 2004. Diante dessa realidade, o acesso aberto é visto por Chan e Costa (2005) como um elemento capaz de provocar mudanças significativas na comunicação científica, uma vez que as ações fornecem condições de acesso às publicações de forma mais equitativa, além de incorporarem os pesquisadores dos países em desenvolvimento ao *global knowledge commons*. Assim, concluem que o acesso aberto tem possibilitado melhorar o fluxo de comunicação entre os países em desenvolvimento (Sul-Sul) e entre estes e os países desenvolvidos (Sul-Norte).

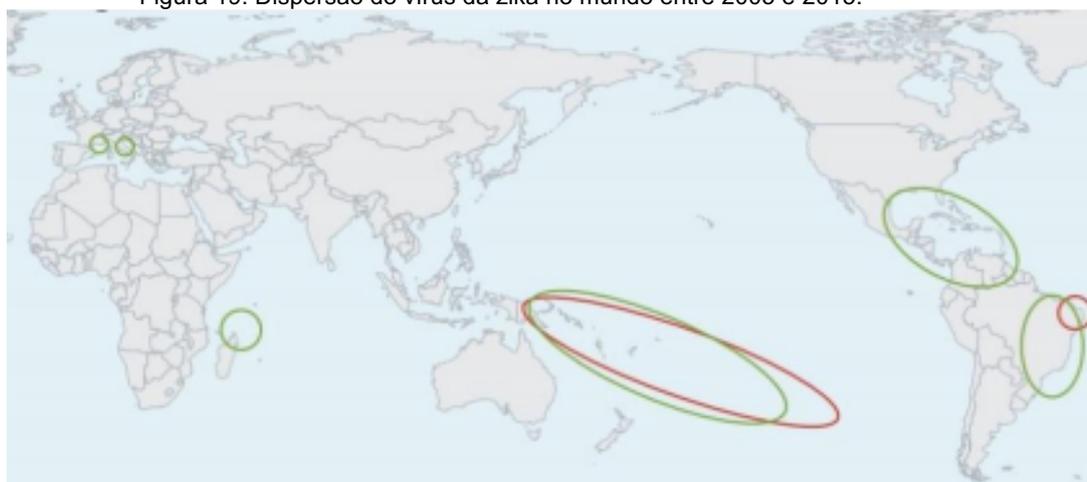
A perspectiva da ciência aberta e da comunicação dos dados de pesquisa, portanto, parece ser especialmente vantajosa em áreas nas quais a produção de conhecimento científico seja restrita e onde exista uma demanda latente por ampliação da visibilidade dos resultados de pesquisa, como é o caso das doenças negligenciadas. A necessidade da prática para a superação de epidemias na saúde pública é mencionada em vários documentos e teve sua expectativa validada pelo acordo de cooperação internacional para a comunicação dos dados de pesquisa sobre as doenças causadas pelo vírus da zika. Portanto, considera-se que a experiência de ciência aberta empregada na comunicação científica sobre o vírus da zika tem potencial para contribuir para a discussão e a implementação de novas práticas científicas sobre outras doenças negligenciadas que afetam a saúde da população e que são de interesse público.

2.3. Contexto social da doença

As doenças causadas pelo vírus da zika, também conhecido como ZIKV, despertaram a atenção das organizações mundiais a partir de 2014, tornando-se realmente visíveis para os governos e suas populações em 2015. Apesar de o estado de alerta sobre o vírus ser recente, desde 1947 ele já era conhecido. Sua primeira aparição foi registrada em um primata na floresta de Zika, localizada em Uganda. A primeira notificação do vírus

em humanos foi feita em 1952, na Uganda e na Tanzânia. Nas décadas subsequentes, foram registradas algumas infecções do vírus principalmente em países da África subsaariana (CHAVES et al., 2015). De acordo com Vasconcelos (2015), a primeira forte epidemia relacionada ao vírus se deu a partir de 2004 em países africanos e asiáticos, chegando em 2013 a terras caribenhas. Vasconcelos (2015) e Zanluca et al. (2015) acreditam que o vírus tenha chegado ao Brasil em meados de 2014, por causa da grande movimentação de estrangeiros no país em virtude da realização da Copa do Mundo FIFA. Apesar da probabilidade especulada pelos autores, Chaves et al. (2015) asseguram que “somente no início do ano de 2015 foram registrados os primeiros casos confirmados de infecção do ZIKV no Brasil” (Figura 19).

Figura 19: Dispersão do vírus da zika no mundo entre 2005 e 2015.



Fonte: (BRASIL, 2016b).

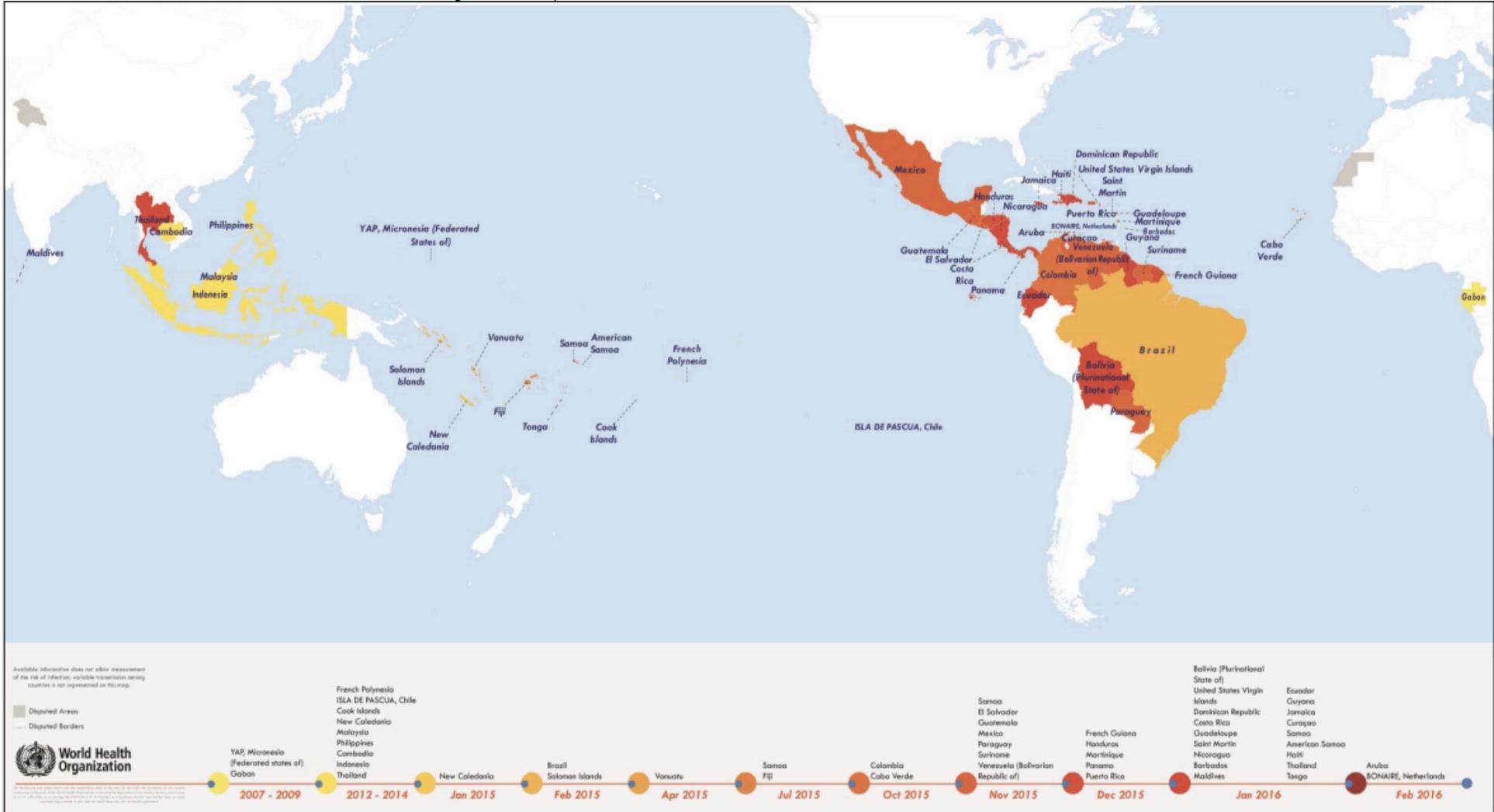
No Brasil, segundo os dados do relatório do Ministério da Saúde sobre microcefalia, ainda não é possível quantificar os casos de infecção do vírus da zika, visto que “cerca de 80% dos casos infectados não irão manifestar sinais ou sintomas da doença e grande parte dos doentes não irá procurar serviços de saúde”. Ademais, são consideradas as limitações atuais de diagnóstico da doença (BRASIL, 2015, p. 16). Com base em projeções internacionais, o Ministério da Saúde estimou um intervalo de 500 mil a 1,5 milhão de casos de vírus da zika no Brasil para 2016 (Tabela 1).

Tabela 1: Projeção de infecções pelo vírus da zika para 2015.

Unidade federada	Estimativa de infecções pelo vírus da zika	
	Limite inferior	Limite superior
Alagoas	4.023	29.066
Amazonas	3.119	34.264
Bahia	19.216	132.274

COSTA, M. **Fatores que influenciam a comunicação de dados de pesquisa sobre o vírus da zika, na perspectiva de pesquisadores.** 2017. Tese (Doutorado em Ciência da Informação) - Universidade de Brasília, Brasília, 2017.

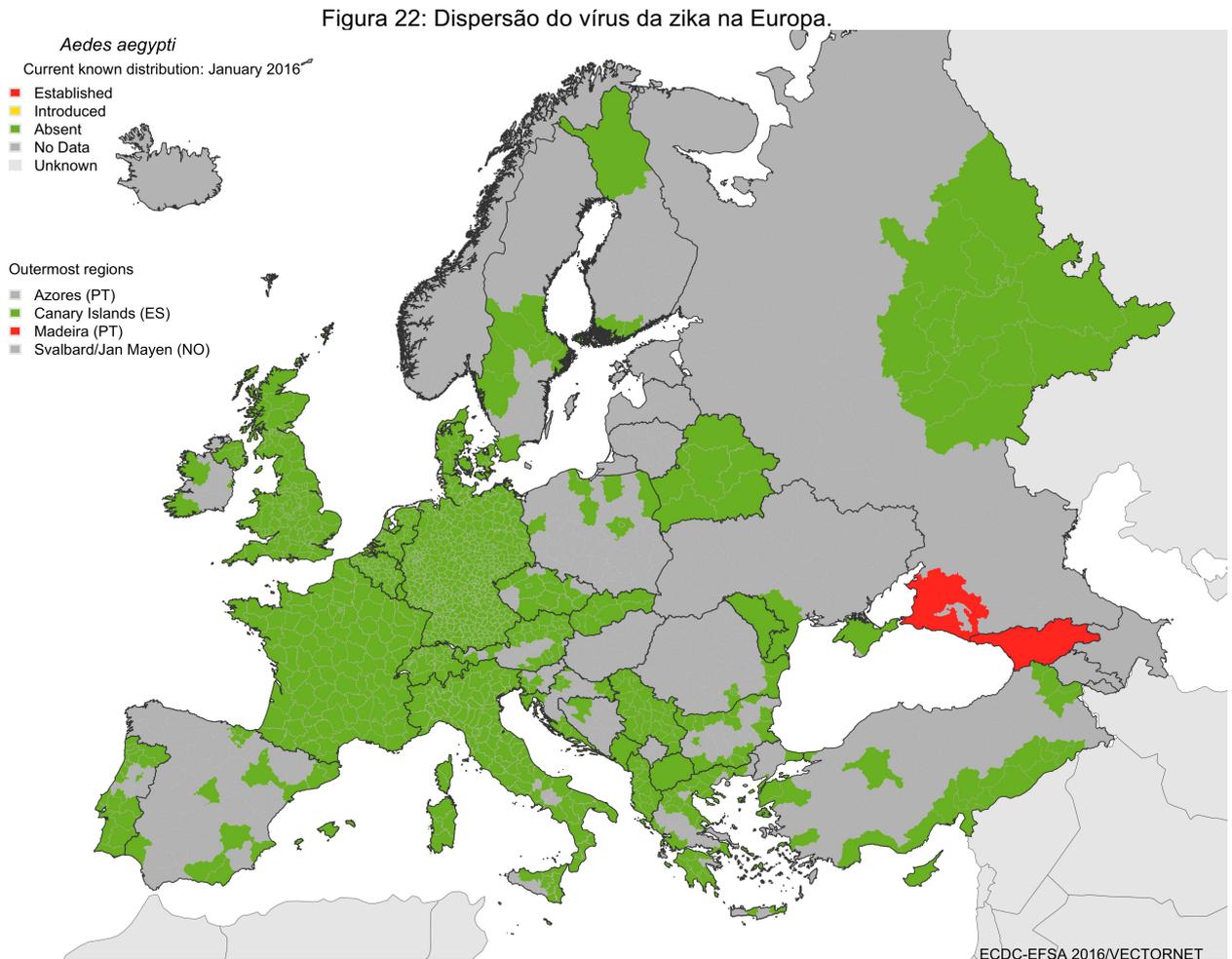
Figura 21: Dispersão do vírus da zika no mundo.



Fonte: WHO (2016a).

COSTA, M. Fatores que influenciam a comunicação de dados de pesquisa sobre o vírus da zika, na perspectiva de pesquisadores. 2017. Tese (Doutorado em Ciência da Informação) - Universidade de Brasília, Brasília, 2017.

O European Centre for Disease Prevention and Control (ECDC) registrou algumas pessoas infectadas pelo vírus da zika em países da Europa, mas apenas no sul da Rússia e na Geórgia do Sul (território britânico) foram localizados focos de transmissão do vírus (Figura 22). Embora o foco de transmissão tenha sido identificado apenas nas duas áreas, a ECDC declarou como regiões de risco todo o sul da Europa, incluindo a Espanha, a França, a Itália e a Grécia (ECDC, 2016).

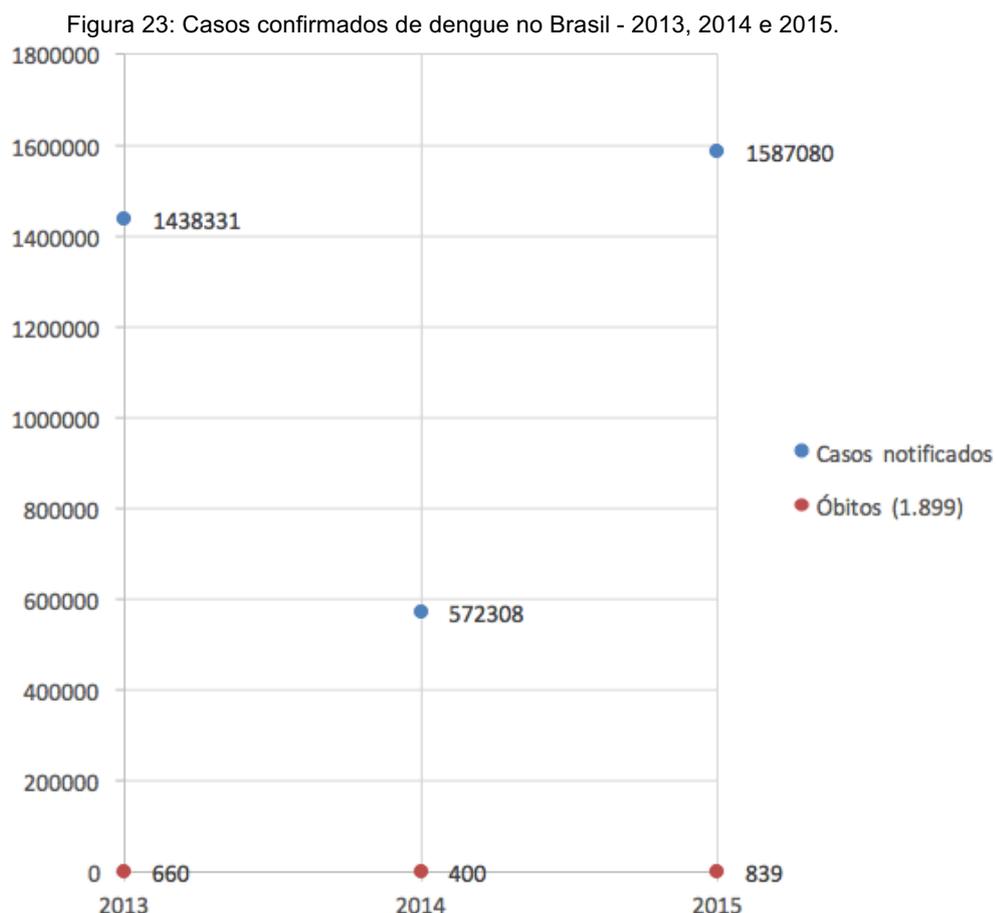


Fonte: ECDC, 2016.

De acordo com Chaves et al. (2015), Zanluca et al. (2015), Vasconcelos (2015) e os documentos oficiais emitidos por Brasil (2016) e WHO (2016), as doenças causadas pelo vírus da zika têm formas de transmissão, sintomas clínicos e possíveis métodos de tratamento semelhantes a outras duas doenças também transmitidas pelo *Aedes aegypti*: a chikungunya e a dengue. A chikungunya também teve suas primeiras manifestações em 1950, na Tanzânia, e chegou ao Brasil em 2014. Em 2015, foram notificados cerca de 20

mil casos da doença no país (BRASIL, 2016a), e os primeiros casos de óbito em decorrência da doença ocorreram no primeiro mês de 2016 (CANCIAN, 2016).

Diferentemente da chikungunya e da zika, a dengue tem pelo menos três décadas de existência no Brasil, já que, em 1981, foi documentada a primeira ocorrência do vírus no país (TEIXEIRA et al., 2001). Segundo Chaves et al. (2015), a dengue está “presente em praticamente 100% do território nacional”. Os dados dos últimos 3 anos do Ministério da Saúde mostram que, em 2015, foram notificados 1,5 milhão de casos no Brasil e 1.900 óbitos nos anos de 2013, 2014 e 2015 (BRASIL, 2014) (BRASIL, 2015c) (Figura 23).



Fonte: Elaboração própria, com dados do Boletim Epidemiológico – Volume 46 – nº 44 – 2015 e Boletim Epidemiológico – Volume 45 – nº 30 – 2014.

O principal ponto em comum das três doenças é o mosquito *Aedes aegypti*, que, segundo Bryant, Holmes e Barrett (2007), é colonizador dos territórios brasileiros há pelo menos 300 anos. A chegada do mosquito ao Brasil, bem como a outros países da América, é atribuída à colonização europeia e ao processo de escravidão de povos africanos pelos

portugueses, que, junto com outras moléstias, trouxe a febre amarela (VASCONCELOS et al., 2004). O antropólogo Darcy Ribeiro também atribuiu o desenvolvimento de epidemias no Brasil também ao processo colonizador do país.

Missões com cerca de 12 mil almas viram-se, em pouco tempo, reduzidas a mil. Nessa situação desesperadora é que ocorrem as epidemias de varíola, de 1562 a 1563, que não atingem os portugueses, mas em três meses matam mais de 30 mil índios e negros. Surge uma nova epidemia na qual morreu mais de um quarto da população indígena sobrevivente. As aldeias, cheias de mortos insepultos, de gente faminta e desesperada, foram abandonadas por muitos índios, que se entregavam aos brancos como escravos, em troca de um punhado de farinha. Por todo o sertão, o desespero grassa também, seja porque as epidemias os atingiram, seja porque os colonos assaltam suas aldeias. Salvos ou induzidos, com toda forma de truques, a ir para a Bahia, onde os escravizam. Dados de Anchieta, em sua "Informação dos primeiros aldeamentos", registram que a população indígena dos arredores da Bahia, avaliada em 80 mil pessoas, se viu reduzida a menos de 10 mil. Às epidemias de varíola, se somou a de febres malignas, completando a destruição (RIBEIRO, 1995, p. 93).

Como narra Vasconcelos (2015), a introdução da febre amarela no Brasil pelos portugueses já foi discutida e apoiada pelos trabalhos de Bryant, Holmes e Barrett (2007), Mutebi e Barrett (2002), Vasconcelos et al. (2004) e Bryant et al. (2005). Nessas abordagens, os pesquisadores identificaram que o vírus da febre amarela está no território brasileiro há pouco mais de 300 anos, coincidindo com o período das grandes navegações e do tráfico de populações africanas que foram escravizadas no Brasil.

Tanto a febre amarela quanto a dengue foram classificadas pelo Ministério da Saúde do Brasil e por outros organismos internacionais de saúde como doenças negligenciadas, que são assim denominadas por receberem pouco ou nenhum investimento para o desenvolvimento de medicamentos (BRASIL, 2010). Segundo Morel (2004), o termo “doenças negligenciadas” é uma evolução política da expressão “doenças tropicais”, que denotava um determinismo geográfico da casualidade das doenças. Na sua visão, essas doenças não são fatalidades das regiões tropicais, mas sim resultado de três tipos de falhas. A primeira delas acontece na ciência, uma vez que o conhecimento científico disponível é insuficiente, implicando a necessidade de promoção de novas pesquisas. A segunda é a falha de mercado, correspondente aos casos em que as vacinas ou os medicamentos existem, mas são inacessíveis em razão do custo, por isso demandam novos meios de produção e financiamento. A terceira, por fim, é a ineficácia da saúde pública, que diz respeito a situações em que as vacinas ou os medicamentos disponíveis são baratos ou até mesmo gratuitos, mas não estão acessíveis à população por causa do planejamento deficiente.

Como elemento característico, essas moléstias ocorrem principalmente em nações em desenvolvimento e nas áreas menos favorecidas desses países. De acordo com Hotez et al. (2006), as condições precárias de vida favorecem a ocorrência das doenças negligenciadas, e as sequelas causadas por essas doenças são condições geradoras de pobreza. Com isso, a doença, que é causada pelas precariedades das condições de vida, contribui ainda mais para o aprofundamento da pobreza em determinadas regiões ou grupos populacionais, acirrando assim o ciclo da pobreza.

A relação estrita que essas doenças têm com a pobreza resulta do baixo investimento para o desenvolvimento de medicamentos capazes de erradicá-las. Na opinião de Nascimento (2012, p. 2), o pouco interesse da indústria farmacêutica sobre tais doenças é devido ao “reduzido potencial de retorno lucrativo para a indústria, uma vez que a população atingida é de baixa renda e presente, em sua maioria, nos países em desenvolvimento”.

De acordo com dados²⁷ apontados por Trouiller e Olliaro (1998), apenas 1% dos 1.393 medicamentos produzidos entre 1975 e 1999 foram destinados ao tratamento de doenças negligenciadas. Em 2006, Chirac e Torreele (2006) atualizaram os dados coletados por Trouiller e Olliaro e abrangeram o período de 2000 a 2004. Seus resultados mostraram que, apesar de terem sido criados mais cinco medicamentos nesse período, a proporção não foi alterada. Apenas 1% dos medicamentos desenvolvidos no período visavam tratar das doenças negligenciadas, que, segundo Pontes (2009), afetam 12% da população mundial. Além de receberem pouco investimento para pesquisa de novos medicamentos, Trouiller e Olliaro (1998) observam que boa parte dos resultados produzidos nesse campo decorre de investimentos de instituições governamentais ou acadêmicas. Moran et al. (2009) identificaram que 69,5% (1,78 bilhão de dólares) dos investimentos para fármacos contra doenças negligenciadas são providos pelo governo, 21,5% (538 milhões de dólares) são oriundos de instituições filantrópicas e apenas 9% (231,8 milhões de dólares) são providos pela indústria farmacêutica.

A discrepância entre o investimento em pesquisa para o desenvolvimento de medicamentos e a ocorrência de doenças graves é recorrente por parte das indústrias farmacêuticas. Conforme Boutayeb (2007), “estima-se que menos de 10% dos recursos de pesquisa biomédica do mundo são dedicados a problemas que lidam com 90% da carga de doença mundial”.

²⁷ Os dados indicados por Trouiller e Olliaro (1998) foram coletados nas seguintes bases de dados: Medline Database, Drugs and Pharmacology Embase, IMS International Data e relatórios da WHO.

Como explicam Garcia et al. (2011), parte significativa dos medicamentos já desenvolvidos para as doenças negligenciadas foi criada em laboratórios oficiais conveniados, com parcerias público-privadas (PPPs), e financiamento do setor público. Os autores asseguram que os governos que investem nessas parcerias visam reduzir o custo da aquisição de medicamentos. Segundo dados expostos pelos autores, o Ministério da Saúde do Brasil articulou nove PPPs para a produção de 24 medicamentos que seriam utilizados pelo SUS. Com isso, o ministério pretendia economizar cerca de 160 milhões de reais com a compra de medicamentos. Do conjunto de medicamentos produzidos pelas PPPs no Brasil, apenas um foi destinado ao tratamento de doenças negligenciadas, a tuberculose.

Na problemática do tema, Castro (2012) analisou a dinâmica do uso de patentes em medicamentos como uma provável razão para o desinteresse da indústria farmacêutica com relação ao investimento para desenvolver novos fármacos para o tratamento das doenças negligenciadas. Como informa o autor, a partir da década de 1980 a legislação sobre patentes favoreceu ainda mais os interesses privados para o desenvolvimento tecnológico, principalmente aqueles relacionados aos fármacos. Segundo Castro (2012, p. 11), “a patente serve para estabelecer os direitos de propriedade e exclusividade de produção e comercialização da criação ou invenção gerada aos seus detentores”. Assim, as indústrias farmacêuticas recebem o direito a um monopólio para a exploração comercial dos medicamentos que produzem. O problema dessa lógica, ainda segundo Castro (2012, p. 12), é que esse regime “não está em consonância com os interesses públicos dos países pobres ou em desenvolvimento”, como no caso das doenças negligenciadas. Logo, os laboratórios não investem no desenvolvimento de fármacos para regiões ou grupos populacionais que não tenham condições de pagar pelos medicamentos. O autor ainda aponta que o acordo internacional²⁸ que garante a dinâmica das patentes permite uma “licença compulsória”, que torna ainda mais desvantajosa para as indústrias a produção de medicamentos para populações pobres, visto que, com essa licença, um país pobre pode obter o direito de produzir uma versão genérica do medicamento e, assim, não gerar lucro para a indústria farmacêutica, como no caso brasileiro, anteriormente citado.

A principal consequência do desinteresse da indústria farmacêutica com relação às doenças negligenciadas é a falta de desenvolvimento de novos medicamentos e a baixa inovação para os fármacos existentes. De acordo com Garcia et al. (2011, p. 37), os medicamentos utilizados para o tratamento de doenças negligenciadas, quando existentes,

²⁸ Trade-Related Aspects of Intellectual Property Rights (TRIPS).

“são aqueles desenvolvidos há mais de 50 anos”. Hotez et al. (2006) acrescentam que, além de tais medicamentos terem sido desenvolvidos há mais de meio século, eles não são eficientes ou seguros.

Como informam Dalcolmo, Andrade e Picon (2007), o primeiro fármaco utilizado para o tratamento da tuberculose data de 1944. Desde então a fórmula sofreu poucas modificações e, mesmo assim, ainda provoca fortes efeitos adversos em razão de seus componentes. Segundo pontos indicados por Vieira e Gomes (2008), os principais problemas relacionados ao uso de fármacos no tratamento da tuberculose são gastrointestinais, cutâneos, articulares e neurológicos.

No Brasil, o tema das doenças negligenciadas foi incluído na Agenda Nacional de Prioridades de Pesquisa em Saúde (ANPPS) e na Política Nacional de Ciência, Tecnologia e Inovação em Saúde (PNCTIS) em 2004, marcando o começo de altos investimentos brasileiros em pesquisa nesse setor, por intermédio do Ministério da Saúde. Para orientar esse processo, foi publicado em 2006 o documento intitulado *Prioridades de Pesquisa em Saúde: Editais Temáticos 2006*, com um caderno dedicado às doenças negligenciadas. Nele, as doenças são classificadas em três categorias, que se referem ao nível de conhecimento e às políticas destinadas a elas.

Diante das problemáticas aqui discutidas, Trouiller e Olliaro (1998) e Garcia et al. (2011) apontam a urgente necessidade de desenvolvimento de novos fármacos, principalmente para aquelas doenças que têm como um dos seus principais aspectos de negligência a falta de pesquisa e desenvolvimento de fármacos.

Nota-se, no contexto delineado, forte relação das doenças causadas pelo vírus da zika com a dengue e a febre amarela, a começar pelo seu vetor de transmissão. No entanto, o vírus da zika tem recebido significativa atenção de pesquisadores, institutos de pesquisas, governo e organizações internacionais, que têm se organizado em um esforço coletivo para a superação da doença. A questão que se propõe aqui é entender por que outras doenças epidêmicas e que afetam grande parte da população não recebem a mesma atenção da ciência. Com isso, espera-se que a análise sobre a comunicação de dados de pesquisa elaborada a partir da experiência das pesquisas sobre o vírus da zika instrumentalize as mudanças necessárias para que as doenças emergentes recebam a atenção devida, tal como o vírus da zika.

2.4. Conclusões da fundamentação teórica

A breve discussão apresentada acerca dos temas pertinentes ao estudo permitiu a conclusão de três elementos. Um deles relaciona-se com o desenvolvimento da concepção de ciência adotada a partir da revolução científica do século XVI e que é reafirmada nas discussões da ciência aberta. Os outros dois elementos, denominados de financiamento e infraestrutura, dizem respeito aos fatores envolvidos no processo de comunicação dos dados de pesquisa e que se mostram relevantes para analisar o contexto de expressão do fenômeno. Ambos os elementos foram sistematizados a partir do conjunto de princípios e recomendações propostos com referência aos cinco documentos que tratam da temática.

O primeiro elemento da fundamentação teórica trata da localização filosófica e histórica da ciência, destacando a importância da comunicação científica para o cumprimento de suas características essenciais para o ideal da ciência moderna. Tal noção foi fundamentada nas características elencadas por Comte-Sponville (2011), onde destaca-se a propriedade da racionalidade. Tal discussão também se fundamentou no relato histórico de Moledo e Olszewicki (2015) que atribuíram o início do pensamento racional na ciência ao pensador Tales de Mileto e suas observações sobre um eclipse solar. Suas observações foram consideradas percussoras do pensamento racional na ciência devido ao uso de dados observacionais coletados por outras pessoas em outras épocas e regiões.

Com isso é reafirmada a centralidade dos processos da comunicação científica para a própria realização da ciência. Conforme apontado anteriormente a noção adotada de comunicação científica baseia-se na sistematização de Garvey e Griffith (1979) dos três macroprocessos, uso, produção e compartilhamento de informação científica enfatizando-se aqui os dados de pesquisa. Na discussão sobre a comunicação científica foi ainda abordado o tema do movimento de acesso aberto, da filosofia aberta e da ciência aberta, como fenômenos atuais e relevantes para a discussão da comunicação científica, e da ciência como um todo, na contemporaneidade. Segundo Boulton (2013), os argumentos motivadores da ciência aberta retomam os princípios da ciência moderna, como forma de realiza-los. Ainda dentro dessa perspectiva de ciência foram destacados dois trabalhos que sistematizaram as categorias da ciência aberta.

Portanto, a discussão acerca da ciência e da comunicação científica permitiram identificar os imperativos da ciência, fundamentais a discussão. Os imperativos são baseados nos princípios da ciência moderna e nas categorias da ciência aberta. Os princípios da ciência moderna são: construção histórica, reconhecimento coletivo,

refutabilidade e ordenação lógica. Já as categorias da ciência aberta representam uma aproximação da proposta feita por Lopez (2015) e pelo projeto FOSTER (2015), onde evidenciam-se que não basta apenas a abertura ou o compartilhamento dos dados para que ele esteja de acordo com a filosofia da ciência aberta. É necessário considerar as metodologias utilizadas para a sua criação, bem como as ferramentas utilizadas, as formas de avaliação, suas políticas e sua relação com as publicações. Com isso, os imperativos da ciência apresentaram-se como uma dimensão de análise do fenômeno estudado.

O segundo aspecto da fundamentação teórica desse estudo tratou especificamente da comunicação dos dados de pesquisa. Primeiramente foram definidos os elementos essenciais para a discussão, tal como o conceito de dado de pesquisa, suas formas de classificação, formas de gestão e ciclos de vida, a partir da discussão de Guibault e Wiebe (2013), OECD (2007), Guidelines (2013), NSF (2014), Corti et al, (2014) e DDI (2013). Ainda como forma de elucidar o contexto foi apresentado um estudo publicado por Costa e Braga (2016) acerca dos repositórios de dados de pesquisa, caracterizando-os.

Posteriormente foi delineada uma narrativa histórica de iniciativas para a comunicação dos dados de pesquisa, que resultou em uma linha do tempo (Costa e Leite, 2017). Aqui vale destacar que foi evidenciado o predomínio de ações europeias e norte-americanas no centro das discussões, bem como das ciências da saúde enquanto área do conhecimento mais presente. Essa parte do estudo também permitiu a identificação de cinco documentos, considerados aqui relevantes para pensar os fundamentos das práticas de comunicação dos dados de pesquisa. Os cinco documentos foram analisados de forma a identificar princípios e recomendações para a comunicação dos dados. Apesar de algumas divergências ideológicas entre os documentos foi possível chegar a um conjunto de 14 elementos que representam as principais recomendações e princípios para o contexto estudado.

Os princípios e recomendações fundamentaram outras duas dimensões de análise do fenômeno. Uma delas é a infraestrutura, composta por três aspectos: legais, tecnológicos e de gestão. Com relação aos aspectos legais foi destacada a conformidade legal das políticas, a proteção da propriedade intelectual, o compartilhamento da mesma licença e o não uso de restrição comercial para os dados de pesquisa. Para os aspectos tecnológicos destaca-se a interoperabilidade e qualidade dos dados. Nos aspectos sobre a gestão foram identificados itens sobre o custo da reprodução dos dados, a unidade e a possibilidade de modificação dos dados, a preservação e o profissionalismo para a gestão dos dados, que é um ponto que traz em especial a ciência da informação para o debate.

A outra dimensão, baseada nos princípios e recomendação enfatiza os aspectos do relacionados ao financiamento da pesquisa, tais como a responsabilidade formal, o equilíbrio de interesses, a prestação de contas e a transparência para ciência.

Além dos aspectos oriundos da fundamentação teórica, outra dimensão foi considerada como necessária para analisar teoricamente o fenômeno da comunicação dos dados de pesquisa. Tal dimensão diz respeito às especificidades relacionadas ao tema objeto das pesquisas que foram analisadas pelo estudo, as pesquisas referentes às doenças causadas pelo vírus da zika. Suas especificidades são resultantes das características do vírus e da relevância que o tema tem recebido nacional e internacionalmente. Suas especificidades foram consideradas como a quarta dimensão de análise dos processos.

3. METODOLOGIA

A metodologia é um modo de pensar e estudar a realidade social, conforme definiram Strauss e Corbin (2008). Seus métodos devem ser sistematicamente organizados para que seja possível executar e compreender a pesquisa. Com essa finalidade, o capítulo apresenta o modelo conceitual, a caracterização da pesquisa do ponto de vista metodológico e os seus procedimentos operacionais. O conjunto desses três elementos compõe a metodologia da pesquisa ora apresentada.

3.1. Modelo conceitual da pesquisa

A pesquisa proposta tem como pressuposto que a comunicação dos dados de pesquisa está inserida dentro de vasto sistema de comunicação científica, caracterizado por três macroprocessos relacionados à produção, ao compartilhamento e ao uso da informação científica, conforme teorizou Garvey (1979). O modo como esses processos acontecem pode variar em função das tecnologias de informação disponíveis e de novos valores e contextos sociais, expressos em forma de fenômeno. O fenômeno que orientou a análise dos processos na contemporaneidade está relacionado à filosofia aberta e representa uma nova maneira de produzir, distribuir e utilizar o conhecimento científico a partir de um contexto denominado ciência aberta. Entre as características da ciência aberta, sobressai a demanda por acesso aos outros produtos resultantes da produção do conhecimento científico, além das publicações revisadas pelos pares. No novo contexto, entende-se que é necessário promover a abertura do conhecimento científico com o uso de tecnologias abertas, metodologias alternativas de avaliação, acesso às publicações científicas e, fundamentalmente, o compromisso de disponibilizar abertamente, para o acesso e o reuso, os dados de pesquisa. Para o modelo conceitual dessa pesquisa, os imperativos da ciência aberta foram considerados como a primeira dimensão interveniente no fenômeno da comunicação dos dados.

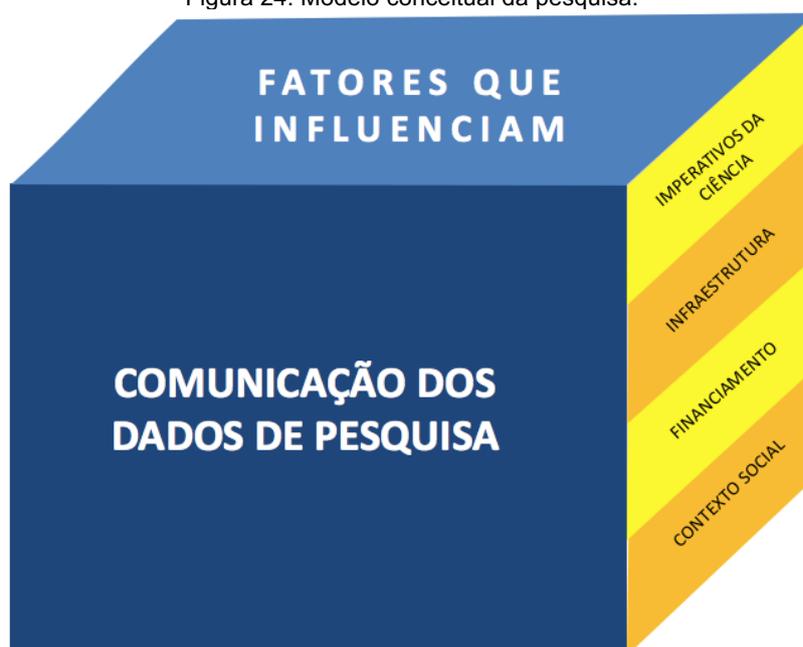
Outras duas dimensões preliminares foram identificadas a partir dos princípios para a comunicação de dados de pesquisa sintetizados no capítulo anterior (Quadro 11). Uma delas foi denominada infraestrutura e considera os aspectos legais, tecnológicos e de gestão dos dados de pesquisa; enquanto a outra, nomeada financiamento, evidencia a influência que os responsáveis pelo financiamento e pelos custos demandados pelos processos podem exercer sobre a comunicação dos dados resultantes da pesquisa.

Além das três dimensões gerais para a análise da comunicação dos dados de pesquisa, considerou-se que o objeto do estudo produz especificidades que influenciam o

fenômeno. A singularidade do objeto é decorrente do contexto social em que ele está inserido. Assim, entende-se que as pesquisas sobre o vírus da zika produzem um contexto social específico que, em razão de sua natureza, intervém no processo de comunicação de seus dados.

Portanto, a teoria embutida no modelo conceitual da pesquisa estabelece que há fatores que influenciam a comunicação dos dados de pesquisa como um fenômeno da comunicação científica, modelado sob quatro amplas dimensões nomeadamente: imperativos da ciência; infraestrutura necessária; financiamento da pesquisa; e contexto da doença (Figura 24).

Figura 24: Modelo conceitual da pesquisa.



Fonte: Elaboração própria.

3.2. Caracterização da pesquisa

Esta pesquisa é de natureza qualitativa, que, segundo Creswell (2010, p. 26), corresponde a “um meio para explorar e entender o significado que os indivíduos ou os grupos atribuem a um problema social ou humano”. Nesse contexto, tornam-se relevantes as particularidades dos sujeitos e temas estudados para elaboração da análise sobre a realidade do problema. Creswell (2014, p. 23) ainda ressalta que “a investigação qualitativa representa um modo legítimo de exploração das ciências sociais e humanas”. De acordo

com Strauss e Corbin (2008, p. 24), esse tipo de pesquisa tem por característica a análise dos dados por meio do “processo não-matemático de interpretação, feito com o objetivo de descobrir conceitos e relações nos dados brutos e de organizar esses conceitos e relações em um esquema exploratório teórico”, diferentemente das análises quantitativas. Portanto, a natureza da pesquisa proposta está em consonância com as discussões metodológicas de sua vasta área do conhecimento, a partir dos procedimentos propostos para a pesquisa qualitativa.

Segundo Creswell (2014), a caracterização da pesquisa inicia-se pela declaração dos pressupostos filosóficos presentes no estudo, especialmente os de natureza qualitativa. O autor conceitua a noção de pressupostos filosóficos como as teorias e paradigmas que formam um conjunto de crenças que orientam a prática da pesquisa. Fundamentalmente o autor defende a declaração dos pressupostos por entender que eles influenciam o modo como o problema de pesquisa é elaborado e, por consequência, a forma como os procedimentos de investigação serão conduzidos. Logo, o autor assegura que a declaração prévia pode ser necessária para a avaliação dos resultados obtidos, uma vez que ela contribui para compreensão das diferenças em relação às abordagens teóricas.

De acordo com a sistematização proposta por Creswell (2014), o pressuposto filosófico desta pesquisa é ontológico. Sob tal categoria, são classificadas as questões relacionadas à natureza da realidade, discutida a partir de múltiplas perspectivas. Tonet (2013) salienta que a abordagem ontológica busca identificar os fatores determinantes que atuam sobre o ser social e que, assim, constituem a sua essência. Logo, a abordagem permite que a teoria sobre a realidade emergja da análise do objeto estudado.

Os pressupostos filosóficos também evidenciam as estruturas interpretativas utilizadas para a discussão do estudo. Creswell (2014) entende que as estruturas são teorias das ciências sociais que contribuem para a estruturação da lente teórica pela qual o problema será analisado. No caso deste estudo, a estrutura utilizada é o construtivismo social, que busca entender a questão da pesquisa a partir dos participantes do processo e seus múltiplos significados impressos para o problema. Apesar de considerar as subjetividades inerentes aos significados identificados, Creswell (2014) salienta que elas são formadas mediante a interação com outros indivíduos e elementos sociais, com normas históricas e culturais que operam na vida do ser social. O autor acrescenta que a proposta de análise construtivista normalmente se manifesta na teoria fundamentada, que se utiliza da perspectiva dos participantes da pesquisa para fundamentar a proposição teórica.

Strauss e Corbin (2008) definem a teoria fundamentada como uma estratégia de pesquisa que permite a proposição de uma teoria a partir da reunião e análise sistemática dos dados de pesquisa; portanto, os autores consideram que a teoria emerge da análise dos dados. Por causa dessa característica, Creswell (2014) considera que a estratégia tem forte orientação positivista.

Entre suas principais contribuições para um projeto de pesquisa, destacam-se seu potencial para oferecer uma explicação sobre determinado fenômeno, bem como sua contribuição para o desenvolvimento teórico de uma área de conhecimento. Para tanto, o uso da estratégia é focado na análise de um processo e de suas categorias teóricas (STRAUSS; CORBIN, 2008).

O principal recurso de coleta de dados para a elaboração de uma teoria fundamentada é o uso de entrevista com amostragem teórica. Essa amostragem contempla sujeitos essenciais para o desenvolvimento do fenômeno estudado. Inicialmente não se define uma quantidade rígida de entrevistados, pois as principais categorias teóricas devem ser exploradas pelos entrevistados até que se alcance um ponto de saturação, ou seja, “nenhum outro dado relevante pareça surgir em relação a uma categoria” (STRAUSS; CORBIN, 2008, p. 205). Os autores consideram que as categorias devem estar bem desenvolvidas em termos de propriedades e dimensões, contemplando suas possíveis variações.

Para Strauss e Corbin (2008), no processo da entrevista é necessário identificar quatro elementos-chave para a exploração do fenômeno:

1. identificação e características do fenômeno central para o processo;
2. condições causais que influenciaram o processo;
3. estratégias empregadas no processo;
4. efeitos e consequências das estratégias empregadas no processo, bem como o contexto em que ele se insere.

A partir dos dados coletados, é possível desenvolver as três fundamentais fases apresentadas por Strauss e Corbin (2008). A primeira trata do processo de conceptualização do estudo, que abrange o processo de descrição, ordenação conceitual e teorização preliminar. A segunda busca identificar as categorias teóricas do processo, realizando-a por meio da microanálise, que permite as codificações abertas, axiais e seletivas.

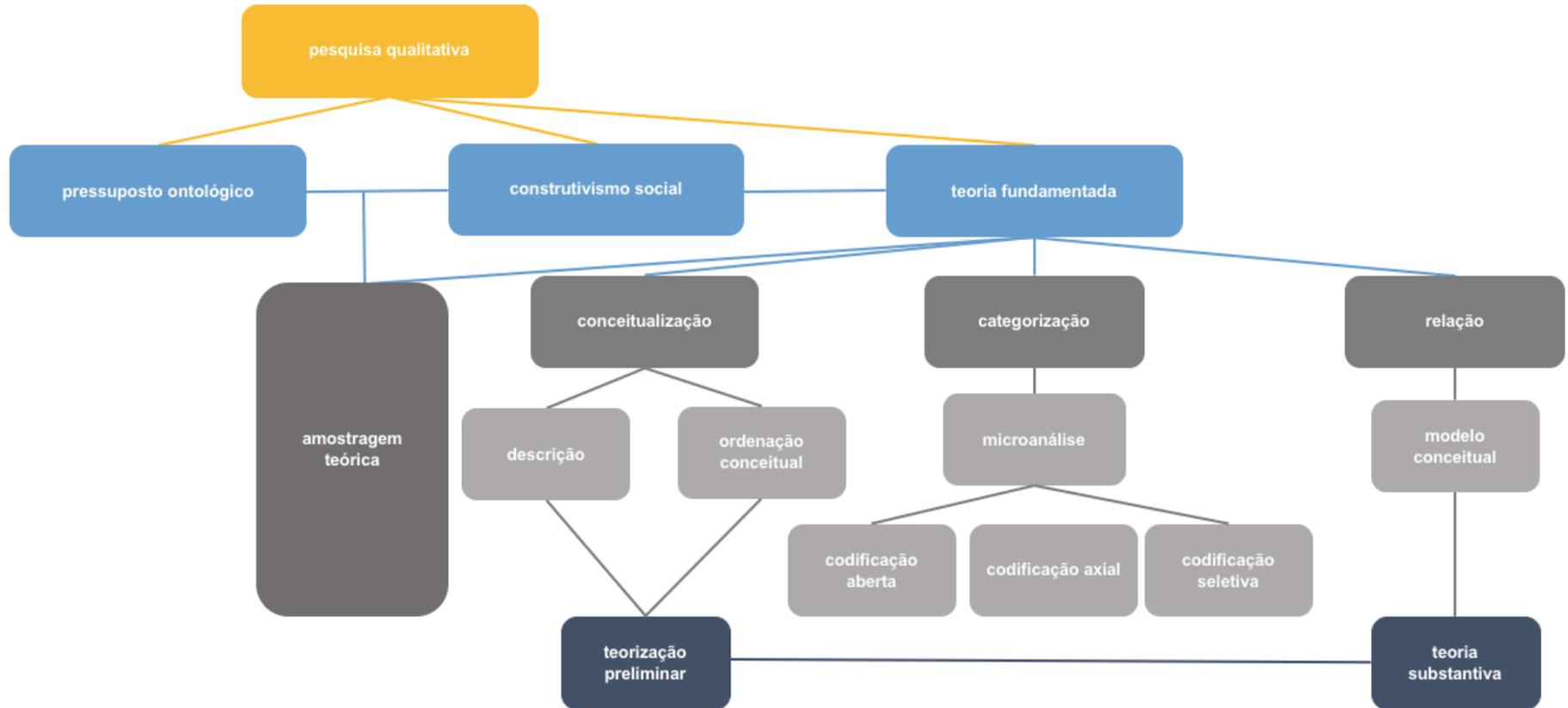
A codificação aberta é o processo analítico por meio do qual os conceitos são identificados e suas propriedades e dimensões são descobertas nos dados. (Strauss e Corbin, 2008, p. 103). Ela permite a redução de grande quantidade de dados em blocos

menores, que podem ser mais bem administrados. Os blocos menores tornam-se a estrutura inicial para construção da teoria.

A codificação axial é o processo por meio do qual se relacionam categorias às suas subcategorias. O processo é “chamado de axial porque ocorre em torno do eixo de uma categoria, associando categorias ao nível de propriedades e dimensões” (Strauss e Corbin, 2008, p. 123). Seu objetivo é reagrupar os dados e relacioná-los às subcategorias, de modo a gerar explicações mais completas sobre o fenômeno estudado.

A codificação seletiva é o processo de integrar e refinar a teoria. A partir da identificação da categoria central, propõe-se a teoria, que posteriormente é refinada. A teoria é acompanhada de um diagrama que ilustra as relações entre as categorias e subcategorias. O refinamento da teoria identifica inconsistência interna e suas variações. Essas codificações permitem estabelecer as relações entre as categorias teóricas e, com isso, propor um modelo condicional e uma teoria de nível substantivo (Figura 25). O modo como cada uma das fases será desenvolvida neste estudo será descrito na próxima seção, que trata dos procedimentos metodológicos.

Figura 25: Caracterização da pesquisa



Fonte: Elaboração própria.

A caracterização da pesquisa permite a identificação de elementos essenciais do construto teórico da pesquisa e que determina a definição dos procedimentos metodológicos, abordados no próximo tópico do capítulo.

3.3. Procedimentos metodológicos

Para viabilizar a proposição do modelo teórico sobre os fatores que exercem influência na comunicação dos dados de pesquisa sobre as doenças causadas pelo vírus da zika, foram definidos quatro objetivos específicos. Os procedimentos para o cumprimento dos objetivos serão detalhados nesta seção. Em relação a cada um deles, foram apontadas as fontes utilizadas, as técnicas de coleta e a análise dos dados.

1. Caracterizar, a partir da literatura científica, fatores que influenciam a comunicação de dados de pesquisa

Inicialmente a pesquisa buscou reconhecer as discussões existentes na ciência da informação sobre a comunicação dos dados de pesquisa. Seu propósito foi identificar como o tema tem sido tratado pela área, para que a análise produzida por esta pesquisa se insira e se beneficie do corpo teórico do campo da ciência da informação, de forma a contribuir e avançar na discussão. Além disso, a etapa teve por finalidade identificar os aspectos que influenciam a comunicação de dados de pesquisa e que já foram discutidos na literatura científica da ciência da informação pertinente ao estudo. Para tanto, buscou-se sistematizar o conjunto dos fatores intervenientes em categorias preliminares.

Para os propósitos declarados, foi selecionada uma base de dados de ampla representatividade na ciência da informação, a Library and Information Science Abstracts (LISA). Segundo sua descrição, a LISA indexa mais de 440 periódicos de 68 países e 20 línguas diferentes no campo da biblioteconomia e da ciência da informação²⁹ (LISA, 2016).

A busca foi realizada nos meses de julho e agosto de 2016. Por meio do argumento de busca [(ti("research data") OR ti("scientific data")) AND sharing], chegou-se a um conjunto de 15 documentos. O critério utilizado para a busca privilegiou os documentos que tratavam do tema do compartilhamento dos dados de pesquisa como assunto principal e que refletiam esse tratamento no título do artigo. Complementarmente foi considerada a variação que o termo "dados de pesquisa" apresenta na literatura, em inglês o termo pode ser tratado como "research data" ou "scientific data" (Figura 26).

²⁹Fonte: <http://www.proquest.com/products-services/lisa-set-c.html>

Figura 26: Busca na base de dados LISA.

The screenshot shows the LISA search interface. At the top, it says "Library & Information Science Abstracts (LISA)" with links for "Basic Search", "Advanced Search", and "About". A search bar contains the query: "(ti("research data") OR ti("scientific data")) AND sharing". Below the search bar, there are options for "Peer reviewed" (checked), "Modify search", "Recent searches", and "Save search/alert". A section for "Related searches" shows "Information sharing". At the bottom, it displays "15 Results" and options for "Cite", "Email", "Print", and "Save".

Os 15 documentos selecionados tiveram seus resumos analisados de forma a certificar a sua pertinência para o estudo proposto. A verificação sucedeu na exclusão de dois documentos, um por representar uma obra monográfica de natureza didática sobre o tema (documento 5) e outro por se tratar de título repetido (documento 11). Treze documentos resultaram da estratégia apresentada (Quadro 12).

Quadro 12: Documentos recuperados na base de dados LISA.

#	Título do documento	Autoria	Seleção
1	Research data sharing: developing a stakeholder-driven model for journal policies	Sturges et al. (2015)	✓
2	Research data management and openness: the role of data sharing in developing institutional policies and practices	Higman e Pinfield (2015)	✓
3	Variability in academic research data management practices	Whitmire, Boock e Sutton (2015)	✓
4	Building a research data management service for the London school of hygiene & tropical medicine	Knight (2015)	✓
5	Managing and sharing research data: a guide to good practice	Gollins (2015)	✗
6	Where Have All the Scientific Data Gone? LIS Perspective on the Data-At-Risk Predicament	Thompson, Robertson e Greenberg (2014)	✓
7	Open access to research data: a track towards scientific collaboration	Melero e Hernandez-San-Miguel (2014)	✓
8	Managing scientific data as public assets: Data sharing practices and policies among full-time government employees	Douglass et al. (2014)	✓
9	Enabling Sharing and Reuse of Scientific Data	Dallmeier-Tiessen et al. (2014)	✓
10	The Conceptual and Practical Ethical Dilemmas of Using Health Discussion Board Posts as Research Data	Bond et al. (2013)	✓
11	The Conceptual and Practical Ethical Dilemmas of Using Health Discussion Board Posts as Research Data	Bond et al. (2013)	✗
12	Upskilling Liaison Librarians for Research Data Management	Cox, Verbaan e Sen (2012)	✓
13	Electronic Scientific Data & Literature Aggregation: A Review for Librarians	Losoff (2009)	✓
14	Sharing research records and research data: findings from a research project in higher education	Childs e McLeod (2004)	✓
15	Compelled disclosure of scientific research data	Gardner (2004)	✓

Fonte: Elaboração própria

Para a análise dos 13 artigos de periódicos selecionados, recorreu-se ao método de Revisão Sistematizada da Literatura (RSL), o qual, segundo Grant e Booth (2009), busca sintetizar evidências de outras pesquisas, podendo incluir ou não pesquisa abrangente e avaliação de qualidade. A síntese de seus resultados é tipicamente em formato de narrativa com acompanhamento tabular (Quadro 13).

Quadro 13: Características da Revisão Sistematizada da Literatura.

Descrição	Pesquisa	Avaliação	Síntese	Análise
Tentativa de incluir elementos de um processo de revisão sistemática, no entanto é mais curta. Tipicamente realizada como tarefa de estudantes de pós-graduação.	Pode ou não incluir pesquisa abrangente.	Pode ou não incluir avaliação de qualidade.	Tipicamente narrativa com acompanhamento tabular.	O que é conhecido, incertezas, por meio da análise de conteúdo; em torno de resultados; limitações da metodologia.

Fonte: Grant e Booth (2009), traduzido e com adaptações.

Para cada artigo selecionado, elaborou-se uma ficha de análise (Quadro 14). O instrumento permitiu sistematizar aspectos sobre a publicação (tal como sua origem e sua data), informações sobre a autoria do artigo (sua origem e filiação institucional), aspectos sobre o conteúdo e sobre o objeto da pesquisa. A investigação sobre o conteúdo dos artigos contemplou os seus objetivos, temas e bases teóricas. Adicionalmente, foram examinadas as metodologias utilizadas pelos estudos. Finalmente, investigou-se o tratamento acerca dos fatores que influenciam a comunicação dos dados de pesquisa e as iniciativas apontadas como facilitadoras do processo.

Quadro 14: Ficha de análise dos artigos de periódicos.

Publicação	Título da publicação	
	Ano de publicação	
	Periódico	
	Origem da publicação	
Autoria	Nome das(os) autoras(es)	
	Filiação da autoria	
	Origem da autoria	
Conteúdo	Tipo de artigo	
	Objetivos da pesquisa	
	Temas tratados	
	Definição dos conceitos	
	Tipo de pesquisa	() Exploratória () Descritiva () Explicativa
	Abordagem metodológica	() Mista () Qualitativa () Quantitativa

Publicação	Título da publicação	
	Ano de publicação	
	Periódico	
	Origem da publicação	
	Métodos	() Estudo de caso () Levantamento
	Técnicas	() Questionário () Entrevista () Pesquisa bibliográfica () Pesquisa documental
	Observações	
	Bases teóricas	
	Autores citados	
	Síntese	
Pesquisa	Perspectiva dos pesquisadores	
	Aspectos que promovem a comunicação dos dados de pesquisa	
	Aspectos que inibem a comunicação dos dados de pesquisa	
	Iniciativas para a comunicação dos dados de pesquisa	
	Observações	

Fonte: Elaboração própria.

A análise sistematizada dos artigos permitiu a elaboração de um conjunto de categorias preliminares acerca do objeto estudado. A ordenação das categorias preliminares foi realizada por meio da rotularização livre do conteúdo e da ordenação das relação conceituais com o auxílio do software Gephi. Essas categorias, segundo Strauss e Corbin (2008), são proposições iniciais, têm caráter transitório e podem ser modificadas ao longo da pesquisa. No entanto, o conjunto preliminar de categorias é um instrumento de orientação para as primeiras entrevistas.

2. Identificar, por meio de entrevistas, a percepção de pesquisadores a respeito dos fatores que influenciam a comunicação dos dados de pesquisa

O cumprimento dos objetivos anteriores possibilitou a sistematização de um conjunto de elementos que representam os fatores que influenciam a comunicação dos dados de pesquisa já discutidos pela literatura científica da ciência da informação. O conjunto alcançado orientou a coleta de dados por meio da entrevista realizada com os pesquisadores envolvidos em estudos sobre o vírus da zika.

Os pesquisadores brasileiros foram considerados o universo dos sujeitos participantes da pesquisa. A identificação dos pesquisadores se deu por meio dos registros de grupos de pesquisa no Diretório de Grupos de Pesquisa no Brasil³⁰ (DGP/CNPq), a partir dos seguintes critérios de busca:

1. Consulta parametrizada. Termo de busca: “aedes aegypti” nos campos “Nome do Grupo de Pesquisa” e “Nome da linha de pesquisa” – 20 documentos;
 - a. Filtro – grupos certificados e grupos atualizados – 10 documentos;
2. Consulta parametrizada. Termo de busca: “zika” nos campos “Nome do Grupo de Pesquisa” e “Nome da linha de pesquisa” – 5 documentos;
3. Adição dos resultados do termo “aedes aegypti” e dos resultados do termo “zika” (resultados 1 e 2) – 15 documentos;
 - a. Exclusão de dois grupos repetidos na primeira e na segunda busca – 13 documentos.

A aplicação da estratégia descrita resultou na identificação de 13 grupos que pesquisam aspectos relacionados diretamente ao vírus da zika e ao *Aedes aegypti*. Portanto, essa foi a amostra intencional selecionada para o universo dos pesquisadores brasileiros. A amostra definida segundo os critérios pertinentes para a pesquisa que se utiliza dos métodos da teoria fundamentada é denominada amostra homogênea teórica. Seu caráter homogêneo representa a singularidade do grupo de sujeitos entrevistados, que aparentemente demonstram o mesmo nível de conhecimento e envolvimento com o fenômeno estudado. Já o caráter teórico da amostra ressalta que os elementos investigados nas entrevistas são categorias teóricas e não buscam representar a pluralidade de opiniões de todos os sujeitos envolvidos com o tema. Portanto, as entrevistas serão necessárias até que cessem novas categorias teóricas. Quando atingido esse ponto, considera-se que a amostra saturou a discussão sobre o processo. A saturação da amostra significa que novas entrevistas não conseguirão identificar novas categorias teóricas, uma vez que as categorias começam a se repetir.

A entrevista foi realizada com o coordenador ou líder dos grupos selecionados e seguiu as estratégias da teoria fundamentada propostas por Strauss e Corbin (2008), relatados na seção anterior. As entrevistas aconteceram com o auxílio de um roteiro semiestruturado, elaborado a partir das categorias preliminares (Apêndice B). Elas aconteceram em duas modalidades: presencial e, em dois casos, virtual. No caso das entrevistas virtuais, foram utilizados softwares de conferências on-line – o Skype e o

³⁰Fonte: <http://lattes.cnpq.br/web/dgp>

Hangout. Todas as entrevistas foram gravadas e seus áudios transcritos. Os pesquisadores entrevistados apresentaram variações em função da área do conhecimento, região, idade e gênero (Quadro 15).

Quadro 15: Caracterização dos pesquisadores entrevistados

Pesquisador	Área do conhecimento	Região	Idade	Gênero
P01	Biologia	Centro-oeste	30-40	Masculino
P02	Biologia	Sudeste	40-50	Feminino
P03	Farmácia	Sudeste	40-50	Feminino
P04	Farmácia	Sudeste	30-40	Feminino
P05	Biologia	Sudeste	50-60	Masculino
P06	Biologia	Sudeste	30-40	Masculino
P07	Biologia	Centro-oeste	40-50	Masculino
P08	Medicina	Sudeste	50-60	Feminino
P09	Medicina	Centro-oeste	70-80	Masculino
P10	Farmácia	Sudeste	30-40	Masculino

Fonte: Elaboração própria

Não foram firmados contratos formais de confidencialidade da pesquisa e consentimento dos pesquisadores. No entanto, foram firmados acordos verbais com os pesquisadores que lhes garantiam privacidade e a omissão de informações que pudessem identifica-los. Portanto, não foram utilizados os nomes dos pesquisadores nos relatos das entrevistas e nos processos de análise e discussão dos dados. A nomeação atribuída aos pesquisadores seguiu a ordem de execução das entrevistas.

3. Analisar os fatores convergentes e divergentes nas perspectivas identificadas sobre a comunicação dos dados de pesquisa sobre as doenças causadas pelo vírus da zika

Os dados obtidos nas entrevistas, bem como os dados sistematizados a partir da literatura científica, permitiram a elaboração de uma teoria preliminar. A teoria e seus dados passaram pelo processo da microanálise, no qual foram feitas as comparações teóricas. No processo de comparação foram identificadas as semelhanças e as diferenças entre os dados e suas dimensões para o processo. Strauss e Corbin (2008) propuseram a execução da comparação teórica a partir de quatro procedimentos:

- 1) elaboração de listagem de propriedades em fontes externas – categorias preliminares;

- 2) verificação das categorias preliminares nos dados e variações emergentes – suas diferenças e semelhanças;
- 3) identificação das dimensões das categorias preliminares a partir dos dados reais;
- 4) explicações mais específicas e densas – evidenciação de categorias não visíveis.

A identificação e descrição das categorias preliminares correspondem a explanação da teoria preliminar. Esta é entendida como ponto de partida para a construção da teoria substantiva. Seu desenvolvimento se dá por meio das comparações teóricas, a partir das codificações dos dados. A primeira delas é a codificação aberta dos dados, que teve por objetivo desenvolver ou aprimorar as categorias em função de suas características e atributos. Com isso, considera-se que o fenômeno foi rotulado pelas categorias e subcategorias que emergiram dos dados.

A etapa seguinte de análise dos dados foi feita por meio da codificação axial, na qual se realiza uma nova montagem das categorias e subcategorias, agora em função de uma categoria central. Nessa etapa, também se iniciou a construção de um paradigma para a relação entre as categorias. Creswell (2014) sistematizou essa etapa em sete passos:

- 1) montagem dos dados de novas maneiras;
- 2) apresentação do paradigma;
- 3) identificação da categoria central;
- 4) exploração das condições causais;
- 5) especificações das estratégias (ações ou interações que resultam no fenômeno central);
- 6) identificação de contexto e das condições intervenientes (condições restritas e amplas que influenciam as estratégias);
- 7) delineamento das consequências (resultados das estratégias).

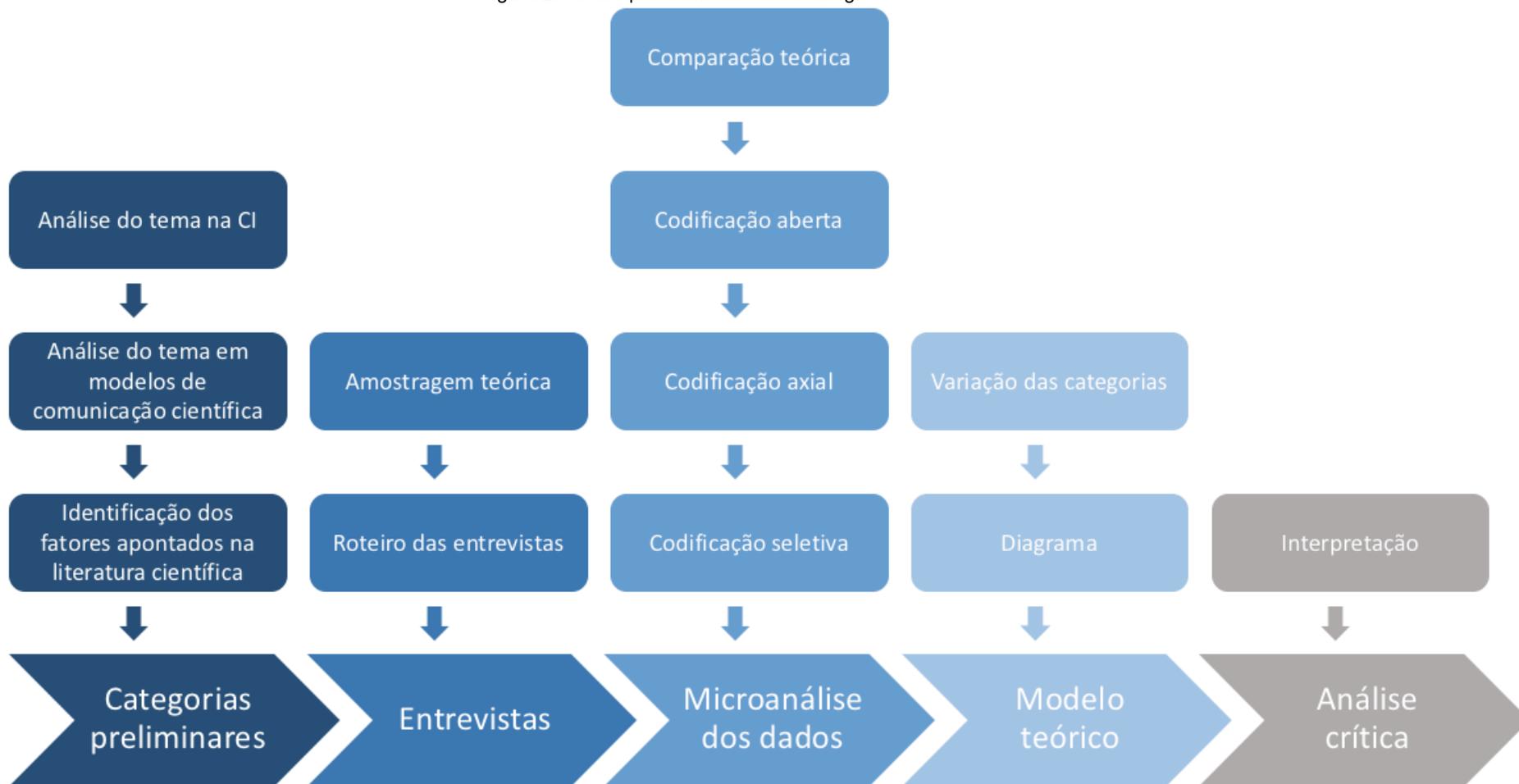
A última etapa da microanálise foi a codificação seletiva, que teve por função refinar a teoria, buscando uma consistência interna e lógica. Nessa etapa, foram aprofundadas as relações entre a categoria central abstrata e as outras categorias e subcategorias, por meio de declarações exploratórias. Segundo a proposta de Strauss e Corbin (2008), a apresentação da teoria já refinada deve contemplar as variações das categorias e ser acompanhada de um diagrama.

4. Examinar as razões pelas quais dados de pesquisas sobre determinadas doenças estão mais propensos a serem comunicados do que os de outras doenças

Os resultados obtidos nas etapas anteriores, bem como o modelo teórico proposto, serviram de instrumentos para sondar as razões pelas quais pesquisas sobre algumas doenças estão mais propensas a terem seus dados comunicados do que outras. A discussão dos resultados foi elaborada por meio da análise crítica dos elementos teóricos e dos dados coletados pela pesquisa.

Todos os macroprocedimentos metodológicos para a execução desta pesquisa estão ilustrados na Figura 27 e detalhados no Quadro 16, em função dos objetivos específicos.

Figura 27: Macroprocedimentos metodológicos.



Fonte: Elaboração própria

Quadro 16: Procedimentos metodológicos da pesquisa

Objetivos específicos	Universo	Amostra	Fonte	Técnica de coleta	Técnica de análise
Caracterizar, a partir da literatura científica da ciência da informação, fatores que influenciam na comunicação dos dados de pesquisa	Literatura científica da ciência da informação	Textos selecionados a partir dos critérios descritos na metodologia	LISA	Pesquisa bibliográfica	Revisão sistematizada da literatura e análise de conteúdo
Identificar, por meio de entrevistas, a percepção de pesquisadores a respeito dos fatores que influenciam a comunicação dos dados de pesquisa	Pesquisadores especialistas do Vírus da zika	Pesquisadores selecionados a partir dos critérios descritos na metodologia	Diretório de Grupos de Pesquisa do CNPq	Entrevista	Codificação aberta, codificação axial e codificação seletiva
Analisar os fatores convergentes e divergentes nas perspectivas identificadas sobre a comunicação dos dados de pesquisa sobre as doenças causadas pelo vírus da zika	Fatores que influenciam o compartilhamento de dados de pesquisa		Resultados anteriores		Análise comparativa
Examinar as razões pelas quais dados de pesquisas sobre determinadas doenças causadas pelo vírus da zika estão mais propensos a serem comunicados no contexto da ciência aberta do que de outras doenças			Resultados anteriores		Análise crítica

Fonte: Elaboração própria

Parte II – Análise dos resultados e discussão

4. APORTE DA LITERATURA SOBRE OS FATORES QUE INFLUENCIAM A COMUNICAÇÃO DOS DADOS DE PESQUISA

Neste capítulo, são relatados os resultados da revisão sistematizada da literatura sobre a comunicação dos dados de pesquisa. A literatura aqui analisada corresponde ao conjunto de 13 documentos resultantes da busca na base de dados LISA, descrita no capítulo anterior (Quadro 12). A revisão sistematizada dos documentos permitiu a identificação de dois elementos relevantes para a pesquisa.

O primeiro elemento trata da caracterização das discussões sobre a comunicação dos dados de pesquisa dentro do campo da ciência da informação. Conforme relatado na fundamentação teórica deste estudo, a ciência da informação é a área do conhecimento que foi criada com o objetivo de estudar os processos relacionados à comunicação científica de modo a aprimorá-los. Com isso, a ciência da informação é uma área do conhecimento que tem por compromisso a discussão de temas relevantes para a comunicação científica, na qual a comunicação dos dados de pesquisa representa uma de suas formas contemporâneas. Para examinar as discussões sobre a comunicação de dados de pesquisa na ciência da informação, levantaram-se dados acerca das publicações dos documentos, suas autorias, objetivos e metodologia de pesquisa.

O segundo elemento diz respeito aos resultados preliminares para o problema desta pesquisa. Nessa etapa, buscou-se identificar o que a literatura da ciência da informação tem apontado como fatores que influenciam a comunicação dos dados de pesquisa. Os resultados obtidos dessa análise serviram de categorias preliminares para a execução da pesquisa por meio da proposta de metodologia da teoria fundamentada nos dados, conforme justificado no capítulo anterior. Segundo Strauss e Corbin (2008), a primeira técnica para o desenvolvimento de uma teoria fundamentada nos dados é o pensamento comparativo. Tal pensamento baseia-se na literatura ou em experiências similares do fenômeno estudado, com o objetivo de estimular a identificação de categorias preliminares.

As categorias preliminares têm por finalidade orientar o roteiro de entrevista, de modo que os sujeitos entrevistados se posicionem a respeito de discussões já existentes sobre a temática, além de evidenciar novas categorias. A estratégia metodológica adotada promoveu o diálogo dos resultados obtidos nas entrevistas com um espectro maior e mais aprofundado de problemáticas discutidas por outros sujeitos relevantes para o contexto da pesquisa. Portanto, a literatura científica da ciência da informação

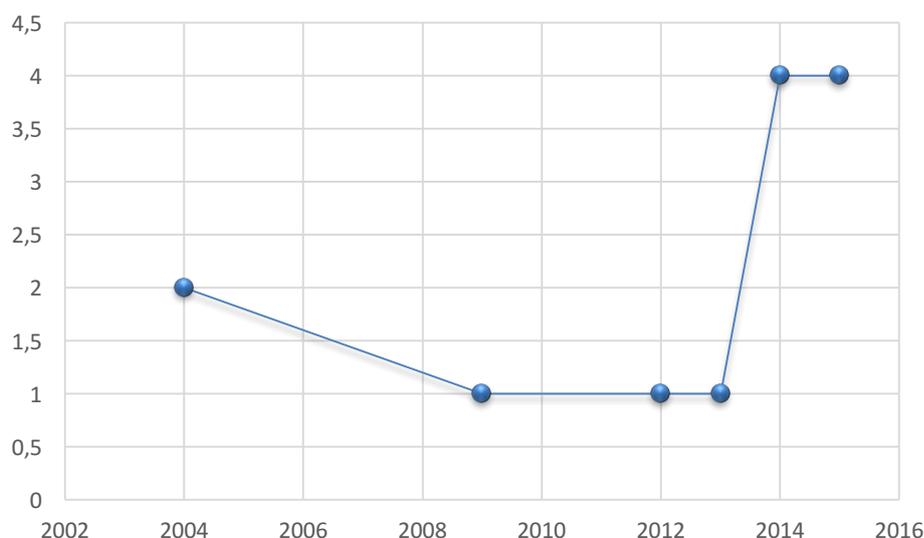
sobre comunicação dos dados de pesquisa apontou as categorias, propriedades e dimensões preliminares que foram consideradas para coletar e discutir os dados obtidos nas entrevistas.

4.1. Caracterização da literatura da ciência da informação sobre a comunicação dos dados de pesquisa

As discussões acerca da comunicação de dados de pesquisa iniciaram-se na década de 1970 com a publicação do modelo UNISIST, conforme apresentado na fundamentação teórica. Nele foi previsto que a interação entre os produtores e usuários de informação científica acontecia por meio de diferentes canais, e um deles seria a comunicação dos dados de pesquisa. Embora o tema tenha surgido precocemente na ciência da informação, ele não se manteve com relevância entre as preocupações da área ao longo dos anos. A temática somente voltou a receber destaque a partir das iniciativas em prol da ciência aberta, no início dos anos 2000.

A tendência ascendente de aparição da temática a partir dos anos 2000 é percebida no conjunto de documentos analisados. Os dois documentos mais antigos datam de 2004, e a maior parte do conjunto de documentos concentra-se entre 2014 e 2015 (Figura 28).

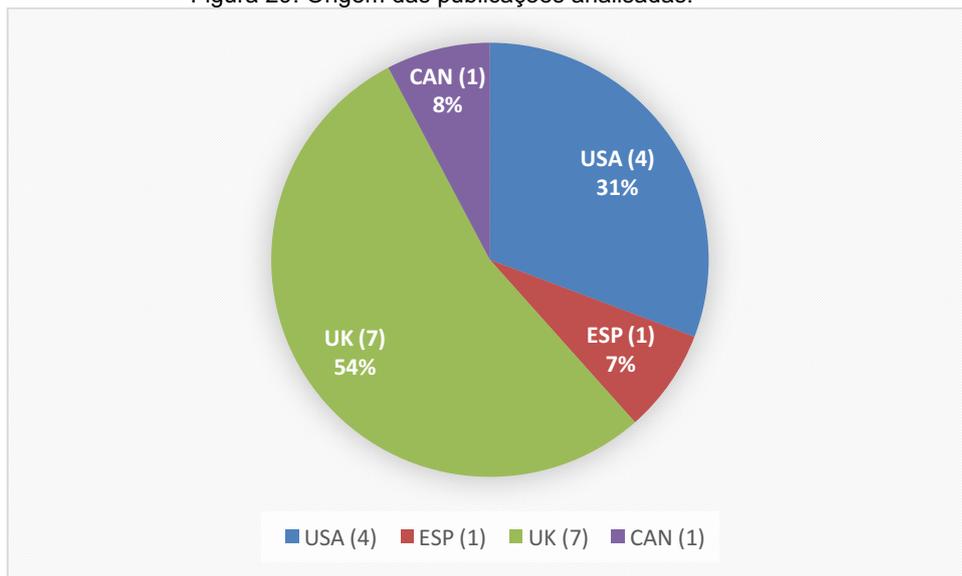
Figura 28: Período das publicações analisadas.



Fonte: Elaboração própria.

Observa-se também que a discussão acerca da comunicação dos dados de pesquisa na ciência da informação não apresenta uma distribuição geográfica uniforme. As publicações em questão são exclusivas de instituições editoras do norte do mundo, com ampla concentração no Reino Unido e nos Estados Unidos (Figura 29). A concentração das publicações nas duas regiões deve-se à predominância de duas editoras, a Wiley e a Taylor & Francis, originárias dos Estados Unidos e do Reino Unido, respectivamente.

Figura 29: Origem das publicações analisadas.



Fonte: Elaboração própria.

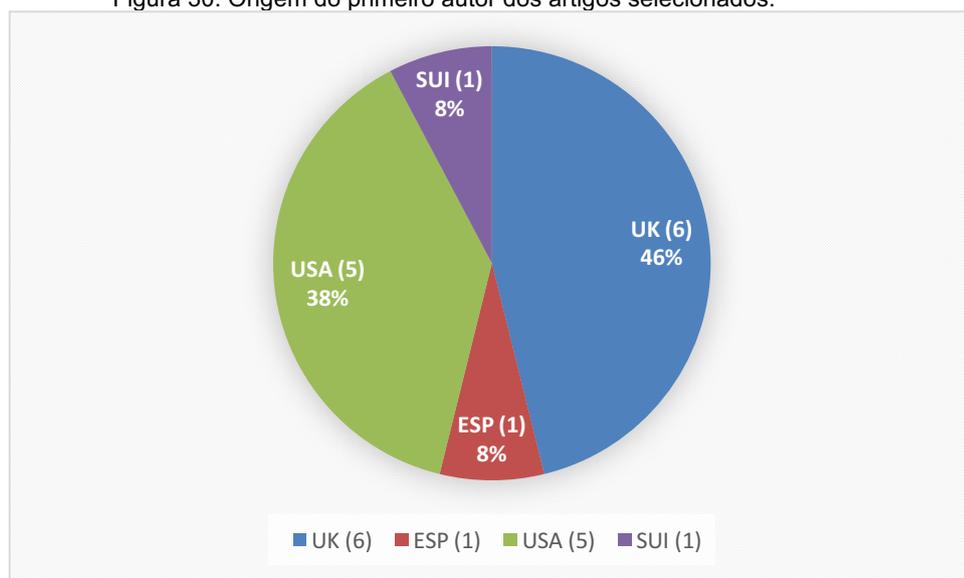
Os dados sobre a concentração das publicações chamam atenção principalmente por causa da natureza das editoras que predominam como publicadoras do tema. As editoras Wiley e Taylor & Francis estão entre as cinco editoras acadêmicas comerciais que mais lucram com a comercialização da comunicação científica no mundo, conforme apresentado no Quadro 1 deste documento.

O antagonismo dos interesses postos demonstram a profundidade da problemática acerca da comunicação científica e da ciência aberta. Editoras comerciais auferem lucros até mesmo de trabalhos que propõem ou defendem ações que visam conferir transparência, abertura e ampla possibilidade de acesso aos frutos da comunicação científica. Não se pretende aqui elaborar uma crítica acerca dos fatores e dos interesses que imbricam tal relação. Apenas se evidencia a predominância de

editoras comerciais na discussão e elas, aparentemente, representam as forças contraditórias das filosofias abertas.

A dominação dos países do norte do mundo no que se refere à procedência das publicações repetiu-se também entre a origem dos autores. Em cada um dos artigos selecionados, buscou-se identificar o lugar de origem do primeiro autor. Como resultado, constatou-se que eles são predominantemente dos Estados Unidos e do Reino Unido (Figura 30). Entre outros elementos, o resultado aponta para a influência do contexto das duas regiões para a discussão do tema.

Figura 30: Origem do primeiro autor dos artigos selecionados.



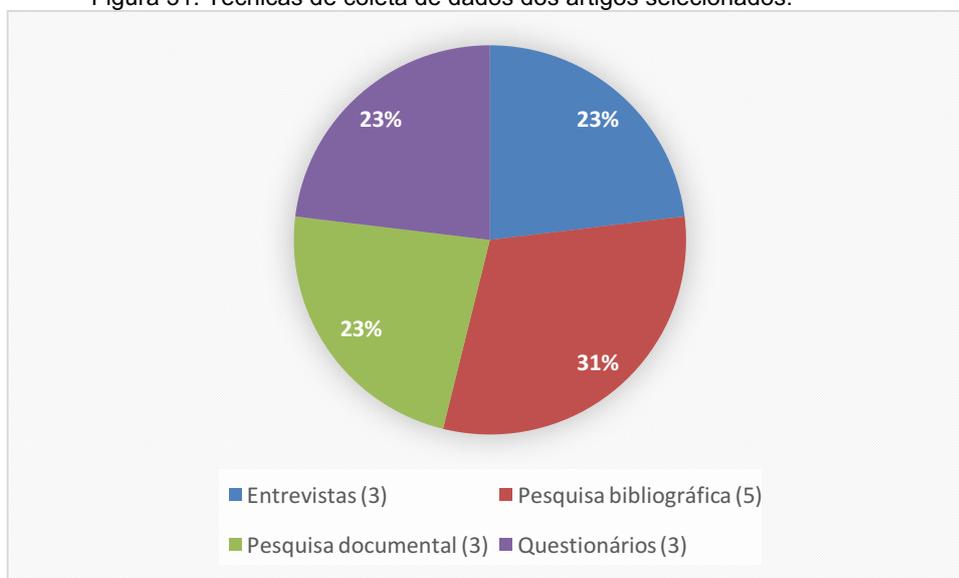
Fonte: Elaboração própria.

Além da procedência das publicações e de seus autores levantou-se, de forma sistematizada, informações sobre a metodologia dos estudos selecionados. Majoritariamente, as pesquisas são de natureza qualitativa. Embora se tenha identificado pesquisas com técnicas relacionadas à abordagem quantitativa, todas as análises foram fundamentadas em princípios qualitativos, mesmo nos estudos de natureza mista.

Entre as técnicas para a coleta dos dados, observou-se uma distribuição uniforme. As principais técnicas identificadas foram pesquisas bibliográficas (31%), entrevistas (23%), questionários (23%) e pesquisas documentais (23%) (Figura 31). Ressalta-se que nenhum dos artigos classificados na categoria “pesquisas bibliográficas” se declaram como dessa natureza. No entanto, eles foram assim

categorizados por não apresentarem dados que amparem suas discussões e resultados, senão a sustentação teórica de revisões bibliográficas.

Figura 31: Técnicas de coleta de dados dos artigos selecionados.



Fonte: Elaboração própria.

Para o desenvolvimento do estudo proposto, todos os artigos selecionados foram examinados com o auxílio da “ficha de análise de artigos” (Quadro 14). Após a coleta dos dados e o exame preliminar, os textos foram classificados segundo sua principal contribuição para a discussão deste estudo. A identificação dos temas abordados nos documentos deu-se por meio da análise de conteúdo de todo o texto. Após o levantamento dos principais tópicos, foram organizados arranjos para a aproximação conceitual entre eles. Os arranjos temáticos foram estruturados em quatro categorias, que serão abordadas nos subitens desta seção.

4.1.1. Atuação de profissionais da informação na comunicação dos dados de pesquisa

Dois artigos analisados trataram especialmente da atuação de profissionais da informação no processo de comunicação dos dados de pesquisa. Um deles foi o artigo de Losoff (2009), que discutiu a integração de sistemas de informação e destacou o papel dos bibliotecários no contexto. O outro é o artigo de Cox, Verbaan e Sen (2012),

que apresentou um projeto de capacitação de bibliotecários para a atuação na comunicação dos dados de pesquisa. A análise de ambos os estudos será apresentada a seguir.

O artigo de Losoff (2009) identifica alguns fatores que influenciam a integração de bases de dados de pesquisa com os periódicos acadêmicos. No estudo, é enfatizado o papel que os pesquisadores e bibliotecários desempenham no processo. A autora considera que a comunicação dos dados de pesquisa tem migrado das bases de dados para os periódicos, que acabam por apresentar as características das bases de dados (repositórios de dados de pesquisa)³¹. Dito isso a autora afirma que a divisão entre os repositórios de dados de pesquisa e os periódicos estaria se tornando cada vez menos evidente. Para sustentar seu argumento, a autora utiliza o estudo de Carlson (2008), que apontou as melhores formas para comunicar dados de pesquisa. A proposta de Carlson (2008) para a temática foi baseada nos resultados de entrevistas realizadas com bibliotecários da Johns Hopkins University, localizada nos Estados Unidos. Portanto, aparentemente o argumento de Losoff (2009) está fundamentado em uma perspectiva de profissionais da ciência da informação, que estão diretamente envolvidos com o desenvolvimento dos repositórios e com o suporte aos periódicos.

A perspectiva de Losoff (2009) ainda é suplementada por outros dois elementos. Um deles é o exemplo do Projeto Genoma Humano, que, de acordo com a autora, colocou os repositórios de dados sobre o genoma humano com mais relevância do que as próprias publicações. O outro elemento é a previsão de evolução da web atual, também conhecida como web sintática, para a web semântica. O progresso da web será marcado pela alteração na forma como a informação é descrita e organizada. A autora cita a proposta de consolidação da web semântica, elaborada por Tim Berners-Lee, reconhecido como o criador da web. Segundo essa proposta, a web semântica criará uma mídia universal de dados, informação e conhecimento intercambiáveis.

Losoff (2009) evidencia uma apreensão com relação à comunicação dos dados de pesquisa que diz respeito ao local onde os dados serão divulgados. Por trás da argumentação da autora, nota-se a preocupação sobre o sistema que será responsável por armazenar e disponibilizar os dados, que conseqüentemente influencia a forma como os dados serão descritos e organizados, principalmente quando considerado o contexto da web semântica.

³¹ O que a autora chamou de bases de dados é tratado aqui como repositórios de dados de pesquisa.

O uso de periódicos para a publicação do conjunto de dados de pesquisa desperta ainda outra questão relevante para o estudo. A questão refere-se aos periódicos de dados, também conhecidos como *data journals*, que são apontados como uma forma de recompensar os autores pela divulgação dos dados.

Candela et al. (2015) realizaram uma investigação na qual identificaram mais de cem periódicos de dados. Considerando que os primeiros periódicos dessa natureza foram criados em 2011, os autores avaliaram que o quantitativo cresce de forma exponencial. Entretanto, apesar das previsões de Candela et al. (2015) e da asseveração de Losoff (2009) pela transição dos repositórios de dados para os periódicos, os primeiros ainda são predominantes. Conforme apresentado na fundamentação teórica deste estudo (pág. 46), no início de 2016 havia mais de 1.500 repositórios de dados de pesquisa registrados no diretório re3data.org.

Cox, Verbaan e Sen (2012) diferentemente de Losoff (2009) não tratam da atuação direta de bibliotecários em sistemas de informação específicos. Os autores analisaram o desenvolvimento do projeto RDMRose, cuja finalidade é promover materiais de aprendizagem sobre gestão de dados de pesquisa para profissionais da informação, especialmente os bibliotecários de instituições de ensino superior. Os autores partem do pressuposto que os bibliotecários ocupam um lugar privilegiado na gestão dos dados de pesquisa e que, portanto, são peças-chave para o processo da comunicação como um todo. A importância de bibliotecários na comunicação de dados de pesquisa está relacionada com a atuação desses profissionais na mediação da informação e no desenvolvimento de sistemas para a comunicação dos dados.

Infere-se da discussão apresentada no artigo que a falta de atuação profissional qualificada no processo de comunicação de dados de pesquisa é um problema para o processo. Isso porque o suporte profissional para a gestão dos dados de pesquisa é apontado como elemento necessário para promover e otimizar as práticas de produção, compartilhamento e uso dos dados de pesquisa por pesquisadores dentro do contexto da ciência aberta. Cox, Verbaan e Sen (2012) enfatizam duas habilidades dos profissionais que são estratégicas para tais finalidades: o conhecimento técnico e a prática para gestão da informação; e o posicionamento privilegiado dos profissionais nas instituições no que tange à política de informação.

4.1.2. Gestão na comunicação dos dados de pesquisa

Uma das vertentes da atuação dos profissionais da informação abrange o conhecimento técnico e as habilidades para a gestão dos dados de pesquisa. Nesse contexto, a gestão envolve os processos de planejamento, manipulação, armazenamento e preservação dos dados de pesquisa, tal como discutido no capítulo 2 (p. 41). No conjunto de artigos analisados, foram identificados três documentos que tratam especialmente de iniciativas que visam à gestão da comunicação dos dados de pesquisa, são eles: Childs e Mcleod (2004), Whitmire, Boock e Sutton (2015) e Thompson, Robertson e Greenberg (2015).

O artigo de Childs e Mcleod (2004) analisa o desenvolvimento do programa Supporting Institutional Records Management, desenvolvido pelo Joint Information Systems Committee (JISC) na Northumbria University, localizada na Inglaterra. O programa é destinado à gestão dos dados de pesquisas desenvolvidas no âmbito da universidade. Para a análise, as autoras entrevistaram 14 pesquisadores com o objetivo de identificar os problemas relacionados à prática de comunicação dos dados de pesquisa.

O estudo, portanto, fornece um conjunto de questões que deveriam ser consideradas no programa. As questões apresentadas pelos pesquisadores entrevistados foram bem variadas, no entanto não foram amplamente discutidas pelas autoras, apenas indicadas no artigo. As problemáticas relevantes identificadas pelo estudo são as seguintes:

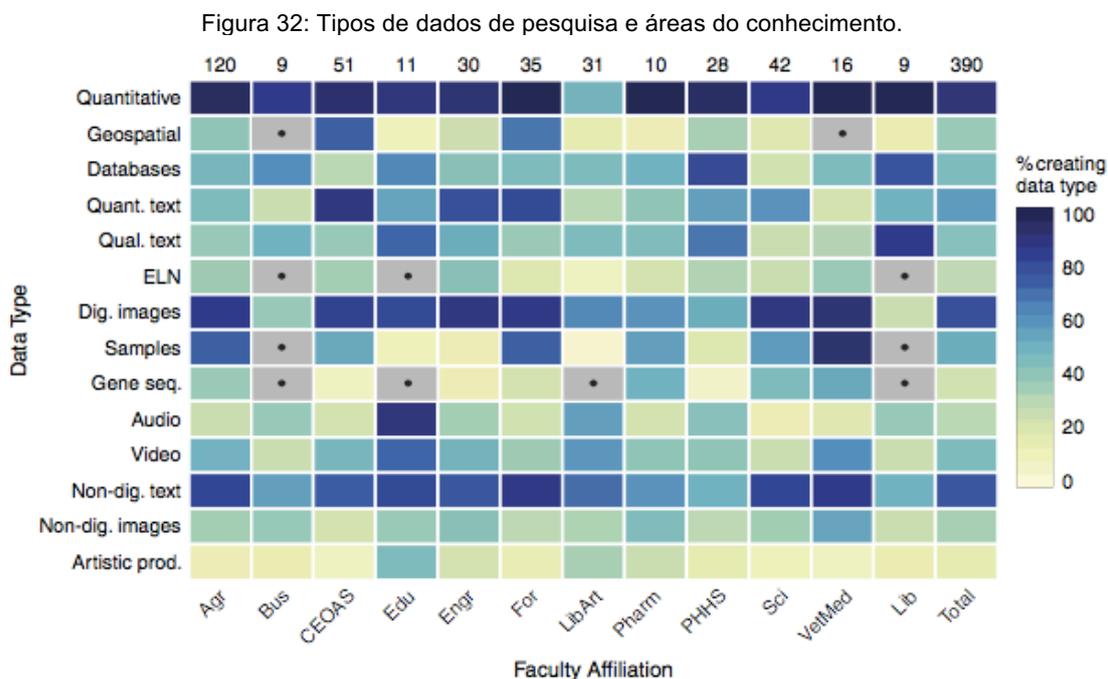
- questões éticas relacionadas à confidencialidade dos dados;
- registro digital dos dados e agrupamento dos *datasets*;
- implicações legais relacionadas ao reuso de dados primários e à retenção dos arquivos de dados primários;
- propósitos do reuso dos dados;
- responsabilidades e esforços para o registro dos dados;
- processos e procedimentos para a comunicação dos dados de pesquisa;
- sistemas tecnológicos envolvidos.

A contribuição do estudo para o programa do JISC foi a proposição de ações para o aperfeiçoamento dos serviços oferecidos aos pesquisadores. De forma geral, as soluções envolvem serviços de orientações aos pesquisadores; elaboração ou

indicação de padrões e políticas; oferta de treinamento aos profissionais envolvidos nos processos de comunicação dos dados; e aprimoramento dos sistemas relacionados aos procedimentos e à infraestrutura tecnológica.

O artigo de Whitmire, Boock e Sutton (2015) também analisou um caso específico de gestão de dados de pesquisa. No entanto, ele teve por objetivo identificar as práticas de pesquisadores no que tange à gestão de dados de pesquisa e às influências que as diferenças disciplinares exercem no tratamento dos dados. O estudo foi realizado em 2013 na Oregon State University (OSU). Os dados foram coletados por meio de questionários eletrônicos enviados aos e-mails dos membros da instituição. Entre os resultados da pesquisa, foram identificados aspectos relacionados aos tipos de dados; ao volume e ao local de armazenamento dos dados produzidos; e aos responsáveis pela gestão dos dados.

O estudo identificou que os padrões de dados mais comuns são os dados quantitativos, as imagens digitais e os textos. Além disso, os autores demonstraram que os resultados variam em função da área do conhecimento. As disciplinas relacionadas às ciências exatas e da terra produzem majoritariamente dados geoespaciais, de natureza quantitativa. Já nas áreas das ciências humanas e das artes prevalecem os dados textuais, de natureza qualitativa. No campo da saúde pública, nomeado com a sigla PHHS, foram predominantes os tipos de dados quantitativos, bases de dados e textos qualitativos (Figura 32).



Fonte: Whitmire; Boock; Sutton (2015, p. 387).

O volume de dados produzidos pelos projetos de pesquisa também apresentou diferentes proporções entre as áreas do conhecimento, variando entre menos de 1 gigabyte até 100 terabytes. No campo da saúde pública, predominou o volume de dados entre 1 e 100 gigabytes. Cerca de 90% de todos os respondentes afirmaram que todos os dados coletados são armazenados em seus computadores ou *drivers* pessoais e são preservados por aproximadamente de 5 anos.

A investigação de Whitmire, Boock e Sutton (2015) constatou que existe uma diferença bem marcada entre as áreas do conhecimento no que diz respeito a responsabilidades com a gestão dos dados de pesquisa. Entre as áreas das ciências exatas, foi predominante a indicação do pessoal de apoio à pesquisa como responsável pela gestão dos dados. Os autores esclarecem que o pessoal de apoio é composto por técnicos de laboratório e assistentes de pesquisa, que não possuem vínculo direto com a universidade e, normalmente, são remunerados por bolsas de pesquisa. Já entre as áreas das ciências humanas e sociais, foi constatado que a responsabilidade da gestão dos dados de pesquisa é exclusiva dos próprios pesquisadores e, em alguns casos, é auxiliada pelos estudantes de pós-graduação envolvidos na pesquisa.

Entre as conclusões apresentadas por Whitmire, Boock e Sutton (2015) para os resultados da investigação, destaca-se a necessidade de melhoramento e ampliação da divulgação dos serviços para a gestão dos dados de pesquisa já oferecidos pela universidade. Os autores ressaltaram que a universidade dispõe de uma central de armazenamento de dados de pesquisa e serviços específicos para a gestão dos dados. No entanto, esses serviços são desconhecidos pelos pesquisadores ou ainda não se mostraram suficientemente atrativos.

O artigo de Thompson, Robertson e Greenberg (2015) tratou especificamente sobre a preservação dos dados de pesquisa e também mostrou variações em função da área do conhecimento. A relevância da temática é justificada pelos autores em decorrência do risco de perda do acesso aos dados de pesquisa. O risco é resultado da fragilidade e da possibilidade de obsolescência das mídias onde os dados de pesquisa são armazenados. Isso significa que, mesmo que os dados estejam devidamente registrados, descritos e armazenados em mídias adequadas, o seu acesso pode ser comprometido ao longo do tempo caso a mídia se deteriore ou as tecnologias

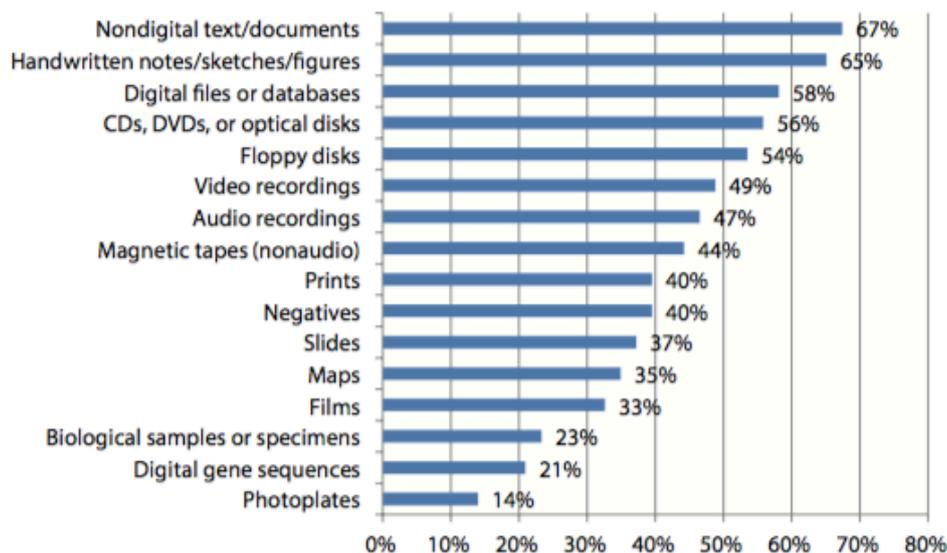
necessárias para acessar a mídia não estejam mais disponíveis em consequência das transformações tecnológicas.

Tendo em vista a problemática do acesso em longo prazo, os autores buscaram identificar as coleções de dados em risco³² e analisar as práticas de preservação dos dados de pesquisas de universidades dos Estados Unidos. Para tanto, foram coletadas informações sobre os tipos e formatos dos dados de pesquisa, além de terem sido elaboradas avaliações sobre o nível do risco das coleções.

As informações foram obtidas por meio de questionários enviados aos bibliotecários e arquivistas, selecionados entre os membros de entidades de classe dos profissionais. De acordo com Thompson, Robertson e Greenberg (2015), são esses profissionais os responsáveis pela curadoria dos dados nas universidades indicadas.

Thompson, Robertson e Greenberg (2015) identificaram que as coleções de dados em risco são majoritariamente de documentos em formato analógico, que incluem documentos de texto (67%) e notas escritas à mão (65%). No entanto, os formatos digitais também aparecem entre as coleções objeto de preocupação do estudo. Entre os formatos digitais, os que receberam maiores indicações de risco foram os arquivos digitais e bases de dados (58%), discos ópticos (56%) e disquetes (54%) (Figura 33).

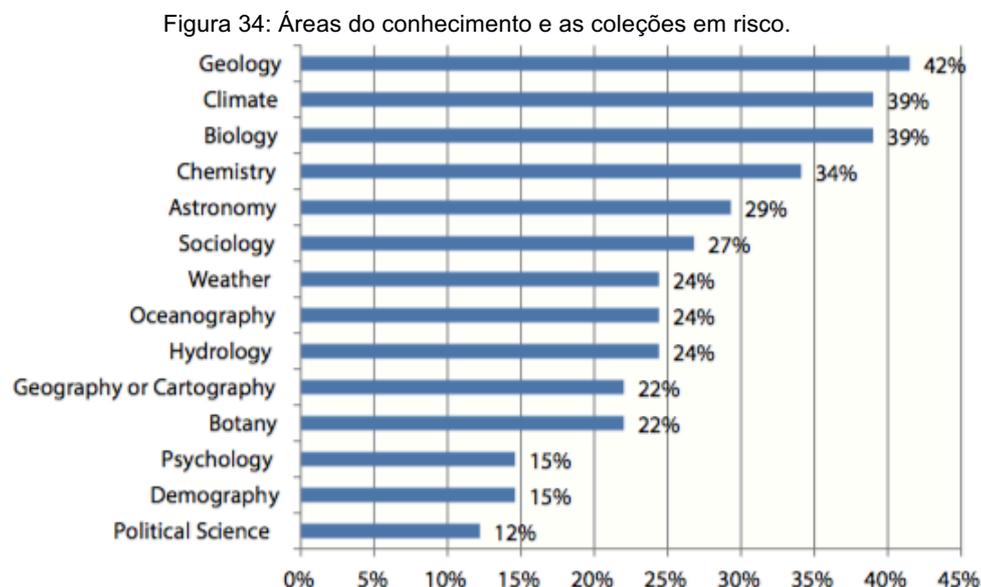
Figura 33: Coleções de dados em risco.



Fonte: Thompson, Robertson e Greenberg (2015, p. 847).

³² Os autores denominaram *collections of at-risk*.

Outro ponto levantado pelo estudo de Thompson, Robertson e Greenberg (2015) foi referente às áreas do conhecimento que possuem coleções de dados em risco. A geologia foi a área que mais recebeu indicações de risco (42%). As ciências da vida, que incluem as ciências da saúde, apareceram em segundo lugar com 39% das indicações (Figura 34).



Fonte: Thompson, Robertson e Greenberg (2015, p. 848)

O estudo buscou também informações acerca dos metadados utilizados para a descrição das coleções. Entre os profissionais que participaram da pesquisa, 64% informaram utilizar padrões de metadados. Os padrões apontados pelos autores são os seguintes: Dublin Core, Federal Geographic Data Committee (FGDC), Space Physics Archive Search and Extract (SPASE), Describing Archives: a Content Standard (DACS), Directory Interchange Format (DIF), e File Information Tool Set (FITS).

Em relação às práticas de preservação, Thompson, Robertson e Greenberg (2015) identificaram que apenas 35% dos respondentes seguem padrões ou políticas para a preservação dos dados. Os padrões considerados pelo estudo incluem o modelo de referência do Open Archival Information System (OAIS), o Preservation Metadata Maintenance Activity (PREMIS) e os requerimentos das agências fomentadoras de pesquisa.

Adicionalmente, a pesquisa levantou os fatores que dificultam a comunicação de dados na perspectiva dos profissionais da informação. Os principais fatores dizem

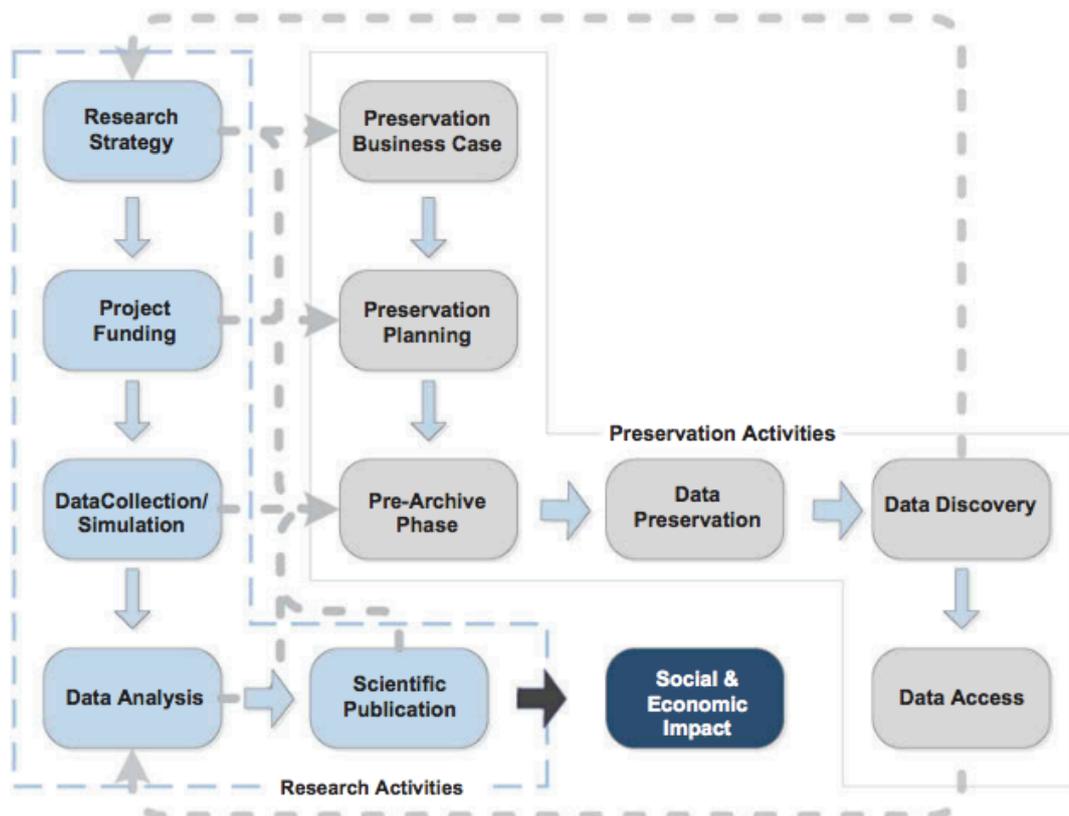
respeito ao tempo envolvido para tornar os dados utilizáveis por outras pesquisas (72%), o acesso aos arquivos armazenados em mídias ou repositórios, (59%) a manutenção da privacidade dos participantes em pesquisas com humanos (35%) e a proteção aos direitos de propriedade intelectual (35%).

Dentre os quatro artigos que trataram da gestão na comunicação dos dados de pesquisa, destaca-se o trabalho de Dallmeier-Tiessen et al. (2014), que consiste na proposição de um modelo que descreve de modo ideal os processos de comunicação de dados de pesquisa, o seu percurso, as barreiras e os facilitadores que determinam o engajamento das partes interessadas (*stakeholder*). Os autores justificam o estudo a partir do contexto de emergência da comunicação dos dados, portanto as políticas e estruturas (denominadas pelos autores de *scientific data e-infrastructure*) que surgirão no contexto precisam ser embasadas em um modelo realista para que possa superar as barreiras existentes.

O modelo proposto pelos autores é uma idealização dos processos, fundamentado em resultados de entrevistas realizadas com pesquisadores, gestores de agências de fomento à pesquisa e editores acadêmicos. As entrevistas foram realizadas entre 2011 e 2012 e abarcou entrevistados de origem europeia, norte-americana e australiana.

Na proposição dos autores, o início de uma pesquisa pode ser promovido ou facilitado pelo acesso a outros dados de pesquisa, que também apresentam sistemas de preservação de dados e devem ser considerados para o plano de preservação (ou gestão dos dados de pesquisa) que será aplicado à pesquisa emergente. O plano de preservação (ou gestão) é (ou deve ser) um requisito para o financiamento da pesquisa, precedendo a fase de coleta de dados. Os dados coletados devem ser arquivados, preservados e preparados para sistemas de descobertas, que os tornarão acessíveis. Essas atividades demandam processos que garantam a interoperabilidade dos dados, tais como clareza e detalhamento na descrição, anotações e informações contextuais. Os autores destacam que aqui é fundamental a atuação de profissionais da informação para o sucesso da fase. A análise dos dados coletados permitirá os resultados da pesquisa, que são comunicados em publicações científicas (Figura 35).

Figura 35: Modelo de comunicação de dados de pesquisa proposto por Dallmeier-Tiessen et al. (2014).



Fonte: Dallmeier-Tiessen et al. (2014, p. 21).

Uma das grandes contribuições de Dallmeier-Tiessen et al. (2014) para esse estudo foi a sistematização dos fatores de influência no processo de comunicação dos dados de pesquisa. As influências foram sistematizadas em função do seu resultado para o processo, portanto foram classificadas em condutores e barreiras da comunicação dos dados de pesquisa.

Os condutores são os fatores que promovem a comunicação dos dados de pesquisa. Eles foram traduzidos pelos autores quanto aos benefícios que geram socialmente para a área acadêmica, para as pesquisas, para as organizações e para os pesquisadores. Os benefícios sociais abrangem aspectos econômicos, educacionais e governamentais. Os benefícios acadêmicos evidenciam as contribuições para a integridade da ciência e o crescimento do entendimento público sobre as pesquisas. Os benefícios para a pesquisa destacam ganhos para quem produz e para quem utiliza dos dados. Para os produtores, é apontado que a comunicação dos dados contribui para a validação dos resultados, bem como para o reconhecimento da produção. Para os usuários, evidenciam-se as facilidades para encontrar novos resultados e tendências,

testar novas teorias com dados passados, evitar repetições, facilitar a descoberta de dados, facilitar o entendimento dos dados por meio da interação entre outras disciplinas e combinar dados para gerar novas informações. Os benefícios organizacionais compreendem a melhora do perfil institucional, a preservação dos dados das publicações, a promoção de novos negócios e a redução do tempo e de custo para produção de novos resultados. Os benefícios para os pesquisadores envolvem a preservação para o acesso posterior, a ampliação da visibilidade e da citação, o crescimento de financiamento de pesquisa, o crescimento do controle dos recursos organizacionais, o impacto socioeconômico, o status e a promoção da carreira e a obtenção de prêmios.

Os autores também previram um conjunto de fatores inibidores da prática da comunicação dos dados de pesquisa, chamados de barreiras, as quais foram classificadas em oito categorias.

A primeira categoria aborda aspectos referentes aos incentivos (ou a falta de incentivos) para os pesquisadores, uma vez que o tratamento dos dados de pesquisa demanda um grande esforço de tempo e recursos dos pesquisadores, e a publicação desse material não recebe o mesmo valor acadêmico que os artigos de periódicos. Segundo Dallmeier-Tiessen et al. (2014), a publicação, a reutilização ou a citação dos dados não conta para avaliação dos pesquisadores. Outra barreira apontada nessa categoria enfatiza a ausência de políticas mandatórias, o que também não estimula os pesquisadores a comunicarem os seus dados. Além disso, os autores destacaram a falta de financiamento para a preservação e a falta de padrões apropriados para a descrição do conjunto de dados.

A segunda categoria refere-se à infraestrutura e preservação dos dados. Sobre esse aspecto, os autores apontam como inibidores da prática da comunicação a falta ou deficiência no que se refere à infraestrutura ofertada aos pesquisadores, além dos altos custos que os processos podem gerar. Sobre a preservação, é apontado que a ausência de garantias da permanência das estruturas pode ocasionar a perda dos dados.

A terceira categoria expõe os aspectos de confiabilidade e usabilidade dos dados. Nela é apontada como barreira a falta de segurança para tratar dados não familiares, principalmente quando considerados dados de áreas interdisciplinares. A falta de clareza sobre a qualidade dos dados e dos requisitos para utilizar os dados também é assinalada como aspecto limitador da comunicação dos dados.

A quarta categoria discute o processo de descoberta ou coleta dos dados. A dificuldade destacada aqui decorre da falta de integração das estruturas e dos programas computacionais utilizados no tratamento padronizado dos dados e sua posterior comunicação. Os autores consideram que a barreira poderia ser superada com um suporte adequado.

A quinta categoria diz respeito ao comportamento dos pesquisadores. Segundo Dallmeier-Tiessen et al. (2014), os pesquisadores que tendem a ter uma atitude defensiva em relação à comunicação de dados apresentam diversos receios, entre os quais se destacam a desqualificação dos dados e as divergências de análise que podem surgir a partir de seus dados.

A sexta categoria versa sobre os aspectos financeiros. A principal preocupação gira em torno dos custos que envolvem o tratamento para a ampla comunicação dos dados de pesquisa. Os autores apontam dois cenários problemáticos para a questão. Em um deles falta financiamento para a infraestrutura e para os sistemas de descoberta dos dados. No outro o financiamento existiu, mas há o risco de não se ter retorno em longo prazo, o que acarreta o fim do financiamento e o abandono da infraestrutura necessária para a manutenção dos dados.

A sétima categoria pondera os aspectos éticos relacionados à pesquisa, entre os quais se destaca a preocupação com a privacidade dos participantes da pesquisa. Como solução para essa demanda, propõe-se o anonimato dos dados. No entanto, Dallmeier-Tiessen et al. (2014) consideram que o custo e a falta de padrões para anonimizar os dados é uma barreira para a superação dos problemas que envolvem aspectos éticos.

A oitava categoria trata da legislação ou regulamentação que envolve os processos para a comunicação dos dados de pesquisa. O tópico é controverso por intermediar dois direitos com forças contraditórias: o direito da propriedade privada e o direito de acesso à informação. A questão ainda é intensificada pelos conflitos entre legislações internacionais, regionais, nacionais, de diferentes países e de diferentes agências de fomento à pesquisa.

Após a apresentação dos benefícios e das barreiras para a comunicação de dados de pesquisa, Dallmeier-Tiessen et al. (2014) ponderam que esses elementos terão diferentes intensidades determinadas por cinco variáveis: o estágio da pesquisa; o envolvimento dos interessados; o contexto específico, no qual se destaca a área do conhecimento; o setor da pesquisa e o contexto geopolítico.

Do mesmo modo, o modelo realçou dois temas pertinentes às discussões sobre a comunicação dos dados de pesquisa e que influenciam o processo. Um tema aborda a relação dos dados com as publicações acadêmicas, e o outro a cultura acadêmica.

Aos editores das publicações acadêmicas foi atribuída a relevância de promoção de infraestrutura e serviços que promovem a comunicação e a descoberta dos dados. Os autores destacaram que grandes editoras como a Elsevier, a Springer e a Wiley têm se comprometido com essa linha de atuação e que tal comportamento visa apoiar um dos princípios da STM Brussels Declaration. O princípio em questão afirma que:

Dados brutos de pesquisa devem ser disponibilizados livremente para todos os pesquisadores. Os editores devem apoiar a postagem pública dos dados brutos das pesquisas publicadas. Conjuntos e subconjuntos de dados devem ser submetidos com o artigo ao periódico e, se possível, disponibilizado livremente para outros acadêmicos (DALLMEIER-TIESSEN et al., 2014, p. 29, tradução nossa).

Por meio de entrevistas realizadas com os editores, o estudo constatou que eles têm interesse em serviços de suporte ao compartilhamento de dados, mas não esperam assumir o papel da preservação em longo prazo. A iniciativa destacada como forma de atuação dos editores foi o desenvolvimento de publicações suplementares para os dados de pesquisa, tais como o Earth System Science Data³³ e o Journal of Open Archeology Data³⁴. Apesar de ter sido apontado pelos editores como uma alternativa para a publicação dos dados, outros participantes da pesquisa afirmaram que as publicações suplementares, via de regra, não apresentam os detalhes necessários para análise, mineração e reuso dos dados. Além disso, apontaram a necessidade da avaliação dos dados em um processo rigoroso e transparente, que os permita serem assegurados em um padrão de qualidade. Adicionalmente discutiram a ausência de padrões comuns de descrição e citação dos dados.

Apesar das fraquezas indicadas, os participantes apontaram iniciativas bem-sucedidas para o horizonte de ações futuras, entre as quais destaca-se os repositórios Dray (Biosciences Data Repository)³⁵ e Datacite³⁶. Ambos os sistemas dispõem de serviço de descoberta dos dados e utilizam o padrão de citação DOI.

Por fim, foram discutidas as questões relacionadas com a cultura acadêmica, que, de acordo com os autores, é governada por normas de comportamento expressas

³³ Fonte: <http://www.earth-system-science-data.net/>

³⁴ Fonte: <http://openarchaeologydata.metajnl.com/>

³⁵ Fonte: <http://datadryad.org/>

³⁶ Fonte: <https://www.datacite.org/>

em códigos e práticas, tais como outras manifestações culturais. Os autores argumentam que existem áreas que demandam mais práticas de colaboração do que outras e, por isso, são mais propícias para a comunicação dos dados de pesquisa. Como exemplo de áreas nessas condições, os autores destacam as ciências biomoleculares, astronomia e ciências das terras. Outras áreas como ciências médicas, química e engenharia podem apresentar comportamento propício para a comunicação, no entanto as práticas são determinadas pelos interesses comerciais envolvidos nas pesquisas. Somados aos aspectos culturais e às forças de interesses nas áreas do conhecimento, a tecnologia foi apontada como um relevante aliado para consolidação de práticas mais colaborativas na ciência, em especial entre os pesquisadores mais jovens. O uso das tecnologias também teria possibilitado a convergência de diversas áreas do conhecimento e, com isso, o crescimento do uso interdisciplinar dos dados de pesquisa.

4.1.3. Políticas abertas para os dados de pesquisa

As políticas consideradas nesta seção abarcam as legislações, as normas, as portarias, os decretos e as orientações originárias de instituições de ensino superior e de pesquisa, bem como de agências governamentais. De forma geral, são documentos que promovem a comunicação das pesquisas que financiam no contexto da ciência aberta (FOSTER, 2015). Além de apontado como estratégico para a atuação dos profissionais da informação na gestão dos dados de pesquisa, as políticas abertas representam uma das dimensões da ciência aberta, segundo o projeto FOSTER (Figura 5). No conjunto dos documentos analisados, foram identificados quatro documentos que envergaram a discussão para o campo das políticas abertas, foram eles: Sturges et al. (2015), Higman e Emmott (2015), Douglass et al. (2014) e Gardner (2004).

Sturges et al. (2015) investigaram as políticas de periódicos acadêmicos que tratam da comunicação de dados das pesquisas que publicam. Os autores partiram do pressuposto que os dados de pesquisa são necessários para conferir transparência ao argumento da pesquisa e promover seu reuso. Portanto, seria coerente que a comunicação dos dados de pesquisa fosse objeto de preocupação dos periódicos acadêmicos, manifestada em suas políticas editoriais.

A investigação focou nos periódicos das áreas de ciências exatas, tecnologia e medicina. A amostra foi composta por 371 periódicos listados no Journal Citation Report

e publicada pela Thomson Reuters em 2011. Identificou-se que 61% da amostra disponibiliza política editorial que faz referência aos dados das pesquisas publicadas. Entretanto, 76% das políticas eram consideradas fracas de acordo com a metodologia proposta por Piwowar e Chapman (2008). A proposta de avaliação classificou as políticas em três categorias: fraca, quando apenas sugere a publicação dos dados; forte, quando requer a publicação dos dados; e inexistente, quando não menciona os dados das pesquisas.

Sturges et al. (2015) estabeleceram uma relação entre as políticas consideradas fortes e os periódicos com maior fator de impacto. Portanto, os autores sugeriram que existe uma tendência para que os periódicos com alto fator de impacto apresentem políticas editoriais para a comunicação dos dados, justificando assim, mais uma vez, a importância de tais políticas para a qualidade da comunicação científica.

Considerando o contexto analisado e as recomendações do Recommendations for Open Access to Research Data in Europe (RECODE)³⁷, os autores desenvolveram um modelo para elaboração de políticas que visam à comunicação dos dados de pesquisa abertamente. Entre os elementos que compõem a proposta, destacam-se os seguintes pontos: requerimento para o depósito dos dados em um repositório adequado; uso de URLs únicas e permanentes; e recomendação para o uso de licenças Creative Commons.

O trabalho de Higman e Emmott (2015) tratou de outra natureza da política e examinou os elementos que conduzem a elaboração de uma política para gestão de dados de pesquisa em instituições de ensino superior. Para tanto foram analisadas 37 políticas de instituições do Reino Unido. De acordo com os autores, tem sido cada vez mais crescente a aplicação desse tipo de política nas instituições de ensino superior e nas agências de fomento.

Aqui é relevante destacar que o Reino Unido (UK) tem um contexto privilegiado na discussão acerca da comunicação dos dados de pesquisa quando comparado a muitos países como o Brasil. Na região, já foram publicados diversos documentos que tratam da questão, inclusive com caráter de legislação para todas as agências de fomento participantes do Research Councils UK (RESEARCH COUNCILS UK, 2011). Portanto, para boa parte dos países ainda não pode ser considerado que a existência de políticas dessa natureza seja um crescente, como sugeriu Higman e Emmott (2015).

³⁷ Fonte: <http://recodeproject.eu/>

Os autores justificam a aplicação de políticas para a comunicação dos dados com base em três argumentos. A possibilidade de verificar os resultados e conferir transparência à ciência é apresentada como o primeiro argumento. O segundo, parte do pressuposto que resultados de pesquisas financiadas com recursos públicos são bens públicos, portanto é errado restringir o acesso aos bens públicos em benefício de interesses privados. O terceiro argumento toca nos interesses pessoais ou institucionais dos pesquisadores e, como os próprios autores definem, é menos altruísta. No argumento, a ampla divulgação dos dados relaciona-se ao aumento da taxa de citação dos artigos, conferindo, assim, maior visibilidade à pesquisa, ao pesquisador e sua instituição. A relação entre comunicação de dados e aumento da citação é baseado em resultados de outros estudos, entre os quais aqui se destaca o publicado por Piwowar e Vision (2013).

A pesquisa desenvolvida por Piwowar e Vision (2013) analisou 10.555 estudos e comparou as taxas de citação dos trabalhos que disponibilizaram seus dados e aqueles que não o fizeram. Como resultado, foi identificado que os estudos que comunicam abertamente seus dados têm entre 5% a 13% a mais de citações do que outros trabalhos sobre o mesmo tema. A análise foi ponderada por um conjunto de variáveis que visou minimizar outras influências sobre as taxas de citação. O estudo também levantou dados sobre o reuso dos dados comunicados. Piwowar e Vision (2013) concluíram que entre 2003 e 2007, 20% dos dados comunicados foram reutilizados. A perspectiva de reuso dos dados é enfatizada pelos autores como um modo de conferir impacto à pesquisa que teve seus dados reutilizados. Isso porque, no contexto analisado, o reuso demanda a adequada citação do trabalho que coletou os dados.

No mesmo sentido em que foram apresentados argumentos favoráveis à comunicação dos dados, Higman e Emmott (2015) apontaram questões que dificultam a prática. O esforço para a reutilização dos dados é apontado como problemático por envolver altos custos. O tema foi discutido pelo trabalho de Borgman (2011), que enfatizou a diversidade de formas e contextos em que os dados podem ser apresentados. Como resultado das especificidades dos dados e dos contextos em que são gerados pode ser difícil ou não aplicável interpretação dos dados fora de seu contexto inicial. Entretanto Higman e Emmott (2015) ponderaram que a problemática dependerá, em grande medida da área do conhecimento, respeitando assim uma das variáveis da análise que se refere às diferenças disciplinares na comunicação dos dados.

O estudo de Douglass et al. (2014) discutiu as políticas de dados do governo e as práticas dos cientistas de agências governamentais. Os autores justificaram o estudo afirmando que o exame das políticas e das práticas é imprescindível para a compreensão do acesso público aos dados de pesquisa coletados por funcionários do governo.

Douglass et al. (2014) demonstraram a relevância do tema para a política dos Estados Unidos por meio de um conjunto de documentos governamentais que se posicionam positivamente em relação ao tema. Entre esses documentos, destaca-se um memorando da Casa Branca, assinado pelo então presidente dos Estados Unidos, Barack Obama. No documento, Obama determina a abertura das informações do governo e daquelas produzidas com recursos públicos. De acordo com o presidente, “a abertura fortalecerá nossa democracia e promoverá eficiência e efetividade em nosso governo” (OBAMA, 2009).

No artigo também são apresentados os resultados de uma pesquisa realizada pelo DataONE³⁸, que é um projeto desenvolvido pela National Science Foundation (NSF), o qual visa garantir a preservação e o acesso aos dados de pesquisa em longo prazo (DATAONE, 2016). Além de um conjunto de serviços de suporte para gestão dos dados de pesquisa, o projeto oferece um robusto motor de busca, que coleta dados em diversas fontes de informação e possibilita diversos recursos para sua recuperação. Em 2013, foi realizado um levantamento, por meio de questionários eletrônicos, no âmbito do projeto DataONE, cuja finalidade foi conhecer as práticas de comunicação de dados realizadas por pesquisadores vinculados ao governo norte-americano. Os pesquisadores participantes alegaram que não publicam seus dados e os tornam disponíveis em ambiente digital por falta de financiamento (42%), falta de tempo (42%), falta de padrões (17%), falta de local para depositar os dados (17%), por não possuírem os direitos necessários para torná-los disponíveis (16%), falta de requisição dos financiadores da pesquisa (13%), por considerarem que os dados não devem estar publicamente disponíveis (8%) e por acreditarem que os dados não serão necessários para outros contextos (7%).

Entre as problemáticas levantadas pelo projeto DataONE, Douglass et al. (2014) exploraram a questão dos metadados e padrões de interoperabilidade. Para esse tema e para os outros abordados nos relatórios, os autores sugerem um conjunto de

³⁸ Fonte: <https://www.dataone.org/>

alternativas que objetivam a superação das questões impeditivas apontadas pelos pesquisadores.

Assim como o trabalho de Douglass et al. (2014), Gardner (2004) tratou das políticas relacionadas a legislação. Em seu trabalho o autor discute os possíveis perigos de abertura dos dados de pesquisas mediante a exigência do Freedom of Information Act (FOIA). O FOIA teve origem em 1999, no congresso estadunidense, promovido pelo senador Richard Shelby. Segundo Gardner (2004), o senador promoveu esse ato, em parte, porque cientistas que testemunharam em seu comitê se recusaram a disponibilizar seus dados de pesquisa. O senador teria considerado a recusa dos pesquisadores descabida e, não tendo encontrado meio legal que os forçasse a apresentar tais dados, propôs o FOIA.

Por meio do ato, pesquisadores eram obrigados a disponibilizar seus dados a qualquer momento e para qualquer demanda. Após críticas dos pesquisadores, foram excluídos da abrangência do FOIA os dados de análises preliminares, rascunhos (*pre-prints*), planejamentos para pesquisas futuras e objetos físicos. No entanto, Gardner (2004) argumenta que, mesmo com as ressalvas, a lei não é suficientemente clara e não representa os interesses dos pesquisadores, uma vez que o ato sujeitaria os pesquisadores a pedidos frívolos e assédios. Gardner afirma que tais pedidos são motivados por interesses antagônicos aos da pesquisa e teriam por finalidade atrasar ou impedir sua realização.

O autor destaca que o assédio que acontece no âmbito legal favorece a supressão de pesquisas, principalmente de temas polêmicos para a indústria farmacêutica. Nas palavras do autor, “o FOIA fornece uma arma para aqueles com motivos estratégicos para silenciar os investigadores científicos”. Tal argumento baseia-se no trabalho de Hailey (2000), que relata ações de empresas farmacêuticas para impedir a publicação de estudos que divulgam resultados adversos sobre seus produtos.

Como alternativa ao ato que favorece o assédio legal aos pesquisadores, Gardner (2004) sugere práticas sistemáticas de comunicação dos dados de pesquisa, após a publicação dos resultados da pesquisa em artigos acadêmicos. Para justificar a necessidade de ampla divulgação dos dados, o autor recorre novamente ao contexto da pesquisa das ciências da saúde, em que apenas os resultados apresentados na publicação podem ser insuficientes para compreender o contexto do paciente e outras variáveis envolvidas.

A diferença que o autor delinea entre o FOIA e a prática sistemática de comunicação dos dados de pesquisa é que, no último caso, os pesquisadores contariam com uma infraestrutura para comunicar seus dados. Os processos relacionados à divulgação dos dados fariam parte do processo da pesquisa de forma geral, portanto estariam no âmbito do financiamento da pesquisa e com responsabilidade compartilhada com as agências fomentadoras de pesquisas e instituições envolvidas. Em síntese, Gardner (2004) defende que a divulgação dos dados deve ser completa, de forma a promover a reutilização dos dados, no entanto não precisa ser exaustiva. A exaustão apontada pelo autor faz referência às demandas que ocorrem no âmbito do FOIA, no qual é necessário produzir documentação verificável para cada dado. Gardner afirma que, além de ser desnecessária para fins científicos, a divulgação exaustiva inviabiliza outros processos necessários ao desenvolvimento da pesquisa.

4.1.4. Ênfase nas ciências da saúde

As discussões acerca dos dados de pesquisa apresentam amplo conjunto de interesses e fatores que influenciam a forma como a comunicação dos dados acontecerá. Os interesses parecem receber um grau a mais de complexidade quando se trata de pesquisas na área de ciências da saúde, conforme se infere da argumentação elaborada por Gardner (2004). Além do artigo de Gardner, apresentado na seção anterior, outros três documentos enfatizaram as especificidades da comunicação dos dados de pesquisa nas ciências da saúde, a saber: Knight (2015), Melero e Hernández-San-Miguel (2014) e Bond et al. (2013).

No conjunto de 13 artigos analisados neste estudo, quatro deles trataram de questões que evidenciam a especificidade das ciências da saúde. Considerando que os artigos analisados foram selecionados em uma base de dados específica da ciência da informação e que os outros nove artigos não trataram de nenhuma área do conhecimento em especial, as ciências da saúde se apresentam como um tema muito relevante para o debate sobre a comunicação dos dados de pesquisa, inclusive para a ciência da informação e seus profissionais.

O artigo mais recente publicado com essa abordagem é de autoria de Knight (2015) e apresenta um estudo de caso sobre um serviço para gestão de dados de pesquisa em uma universidade inglesa. Sua finalidade é analisar a performance de tal

serviço de acordo com as necessidades dos usuários reais e potenciais, a fim de aprimorá-lo.

Um aspecto de especial interesse para a discussão proposta aqui é o protagonismo da London School of Hygiene & Tropical Medicine, instituto da University of London, para a promoção do serviço. De acordo com Knight (2015), a oferta de um serviço institucional para gestão dos dados de pesquisa iniciou em 2002 para toda a University of London. No entanto, em 2009 a reformulação do serviço foi proposta por um professor da London School of Hygiene & Tropical Medicine.

Partindo do pressuposto do trabalho, infere-se que Knight (2015) considera dois elementos como promotores para a comunicação dos dados de pesquisa: o contexto das pesquisas em saúde pública e a gestão dos dados de pesquisa. De acordo com a autora, pesquisadores dessa temática produzem uma grande quantidade de dados que devem ser armazenados e geridos adequadamente para garantir o desenvolvimento de novas pesquisas. A demanda por gestão de dados, intrínseca às pesquisas de saúde pública, perceberia a oferta de serviços com essa finalidade como uma oportunidade para a comunicação aberta dos dados de pesquisa.

Com a finalidade de entender as necessidades dos usuários potenciais do serviço, o estudo levantou os principais problemas relacionados à comunicação dos dados de pesquisa na perspectiva dos pesquisadores da London School of Hygiene & Tropical Medicine. O principal problema, apontado por 100% dos respondentes, relaciona-se com as “incertezas nas práticas relacionadas com o arquivamento dos dados”. Os problemas medianos para o contexto desdobram-se entre: acordo dos participantes para o compartilhamento dos dados (60%), incerteza sobre o padrão que deve ser aplicado à documentação (58%), questões de segurança (53%), falta de espaço para o armazenamento na rede da faculdade (51%), incerteza sobre as ferramentas de softwares a serem utilizados (46%) e questões sobre a produção do plano de gestão dos dados (43%). Outros três problemas foram apontados por um terço dos participantes da pesquisa, são eles: questões de interoperabilidade (31%), velocidade do acesso aos dados (30%) e incerteza sobre o formato de arquivo a ser utilizado (30%). Adicionalmente, cerca de 40% dos participantes ainda afirmaram identificar outros problemas para a comunicação dos dados de pesquisa que não estavam entre as opções de resposta.

A particularidade das ciências da saúde no que tange à comunicação dos dados de pesquisa também foi destacado no artigo de Melero e Hernández-San-Miguel (2014).

O trabalho teve por objetivo apresentar algumas iniciativas e recomendações, em diversas áreas do conhecimento, sobre como compartilhar, gerenciar e promover o acesso aberto aos dados de pesquisa. No entanto, ao discutir aspectos relacionados às políticas de acesso aberto, foi considerado que as áreas das ciências da saúde são especialmente mais envolvidas com a comunicação dos dados de pesquisa.

A avaliação de Melero e Hernández-San-Miguel (2014) sobre esse tópico foi baseada nas informações coletadas em dois diretórios internacionais de políticas de informação, o Sherpa/Juliet³⁹ e o Melibea⁴⁰. Em ambas as fontes de informação, constatou-se que existem mais políticas de acesso aberto relacionadas a universidades e institutos de pesquisa, no entanto esses documentos geralmente não tratam dos dados de pesquisa. Os autores observaram que são as agências fomentadoras de pesquisa que mais demandam planos de gestão e arquivamento dos dados relacionados às pesquisas que financiam. Além disso, constatou-se que, entre as agências fomentadoras com esse perfil, predominam aquelas relacionadas ao campo das ciências da saúde. Os autores consideram que esse cenário foi influenciado pela iniciativa promovida pela agência Wellcome Trust em 2011.

A ação da agência Wellcome Trust consistiu na reunião de entidades de todo o mundo que financiam projetos no campo das ciências da saúde para a consolidação de um compromisso com a comunicação dos dados de pesquisa. Na ocasião, as entidades emitiram um manifesto no qual se comprometeram a comunicar os dados das pesquisas que financiam. No manifesto, apresenta-se o argumento de que a comunicação dos dados de pesquisa proporciona melhor aproveitamento de recursos, traz benefícios para a qualidade da ciência e acelera o progresso da saúde pública (Wellcome Trust, 2011).

No mesmo ano em que o manifesto foi divulgado, Walport e Brest (2011) publicaram um artigo com o mesmo nome do documento, no qual apontam os benefícios da iniciativa. Os autores são gestores da Wellcome Trust e da Hewlett Foundation e aparentemente falam em nome das agências. Os autores afirmam que “como [agências] financiadoras de pesquisa em saúde pública, nós devemos garantir que a produção científica seja usada para maximizar o conhecimento e os potenciais benefícios para a saúde” (Walport e Brest, 2011, p. 537. Tradução nossa). O argumento dos autores está focado na prática de comunicação dos dados como um meio de maximizar os conhecimentos científicos e melhorar a saúde pública. Apesar dos benefícios

³⁹ Fonte: <http://www.sherpa.ac.uk/juliet/index.php>

⁴⁰ Fonte: <http://www.accesoabierto.net/politicas/>

proporcionados pela prática da comunicação, os autores consideram que, em muitos contextos, ainda não existe infraestrutura adequada, nem padrões técnicos e incentivos necessários para a comunicação dos dados de pesquisa. Adicionalmente, apontam a resistência dos próprios pesquisadores em realizarem a comunicação de seus dados. O comportamento dos pesquisadores em não comunicar seus dados deve-se a intenção de “maximizar seus registros de publicações à custa da mais ampla possibilidade de uso dos dados” (Walport e Brest, 2011, p. 538. Tradução nossa).

Walport e Brest (2011) consideram que a discussão do tema sobre a comunicação dos dados de pesquisa suscita três preocupações principais entre os pesquisadores. A primeira delas está relacionada aos pesquisadores que dispõem de poucos recursos para a análise dos dados e, mesmo tendo feito um trabalho crucial para a coleta dos dados, teriam seus esforços apropriados por pesquisadores com melhores recursos, habilidades e ferramentas para analisar os dados de forma mais rápida. A segunda questão evidencia que a comunicação dos dados envolve custo, tempo e dinheiro, portanto espera-se algum retorno do investimento. Para cada uma das três preocupações, os autores propõem soluções. A terceira diz respeito ao receio que os pesquisadores têm de comunicar amplamente os seus dados e, com isso, ferir os direitos de privacidade dos participantes que forneceram os dados.

Os autores propõem que os direitos e as responsabilidades entre os pesquisadores que geram e que usam os dados sejam equilibrados de forma mais justa. Com isso os dados não seriam disponibilizados para “uso livre”⁴¹ e sim para o “uso justo”⁴². O uso justo também resolveria as questões relacionadas ao retorno dos investimentos na comunicação dos dados, porque, de forma geral, essas pesquisas têm por finalidade o melhoramento da saúde pública global. Sobre a terceira questão, referente à privacidade dos participantes, Walport e Brest (2011) enfatizam que os dados só devem ser comunicados se estiverem adequadamente registrados sob os termos em que foram providos. Os autores consideram que a privacidade dos participantes é soberana e pode ser alcançada por meio do anonimato dos dados. No entanto, os autores apontam que os dados de populações coletados para pesquisas em ciências da saúde têm por princípio ético, que seu uso seja maximizado.

⁴¹ O termo original utilizado pelos autores é *free trade*.

⁴² O termo original utilizado pelos autores é *fair trade*.

O artigo de Walport e Brest (2011) foi publicado no periódico *The Lancet*⁴³, número 377. No seu número seguinte, 378, foi publicado um artigo de Sankoh e Ljsselmuiden (2011) em resposta aos argumentos apresentados por Walport e Brest (2011)⁴⁴. Sankoh e Ljsselmuiden são pesquisadores africanos, de Gana e África do Sul, respectivamente, e apresentaram em seu artigo um conjunto de respostas críticas às proposições de solução das problemáticas que envolvem a comunicação dos dados de pesquisa. Suas críticas também foram sistematizadas em três grandes questões, que fazem referência aos argumentos de Walport e Brest (2011).

Sankoh e Ljsselmuiden (2011) iniciam afirmando que o uso justo deve garantir que todos os pesquisadores tenham iguais condições de infraestrutura e recursos para analisar os dados. Somente essa paridade em relação à distribuição dos meios e às capacidades de análise dos dados poderia conferir justiça ao uso dos dados coletados. O segundo ponto trata do retorno dos investimentos. Sankoh e Ljsselmuiden (2011) questionam se o resultado esperado no financiamento de pesquisas na área das ciências da saúde seja o melhoramento da saúde pública global. Isso porque já existe muito conhecimento científico produzido e comunicado que não é aplicado para o melhoramento da saúde da população de países de baixa e média renda. O terceiro ponto não apresenta uma crítica aos argumentos de Walport e Brest (2011), mas faz uma provocação para que seja mais bem definido o que compreende a classificação de dados da população coletados para pesquisas em ciências da saúde.

O artigo de Melero e Hernández-San-Miguel (2014) não apresenta a discussão de Walport e Brest (2011) e Sankoh e Ljsselmuiden (2011). No entanto, ele suscita a aparição da problemática ao apontar o manifesto da Wellcome Trust (2011). Portanto, os argumentos que emergem desse debate são relevantes para o contexto dessa pesquisa, em especial por tratar diretamente da comunicação dos dados de pesquisa no campo da saúde pública.

Já o trabalho de Bond et al. (2013) oferece insumos para as problemáticas acerca da privacidade dos participantes de pesquisa nas áreas das ciências da saúde. Os autores levantaram a opinião de um grupo de pacientes com diabetes, cujos dados pessoais disponibilizados em ambientes on-line foram utilizados por pesquisas. O

⁴³ Fonte: <http://www.thelancet.com/>

⁴⁴ O artigo de Walport e Brest (2011) foi intitulado *Sharing research data to improve public health*, e o artigo de Sankoh et al. (2011) foi intitulado *Sharing research data to improve public health: a perspective from the global south*.

objetivo era compreender se e como os pacientes entendem que suas informações devem ser usadas por pesquisadores no campo das ciências da saúde.

Com vistas a esse objetivo, Bond et al. (2013) entrevistaram 26 pacientes. De modo geral, os entrevistados concordaram que suas informações estavam em domínio público, por estarem amplamente disponíveis na internet. No entanto, não houve consenso sobre a necessidade de serem comunicados sobre a utilização de suas informações pessoais em pesquisas. Parte dos entrevistados considera que é necessário que o uso de suas informações pessoais em pesquisas seja previamente autorizado. A outra parte dos entrevistados declarou que a autorização não é necessária, pois suas informações contribuirão para as pesquisas que buscam melhorar a qualidade de vida das pessoas que vivem com diabetes e até mesmo para superar a doença. Além da divergência de opiniões, o estudo de Bond et al. (2013) revelou que o grupo não tinha conhecimento de que as informações que postavam em uma rede social era utilizada para fins de pesquisa.

A análise textual de cada um dos artigos permitiu a identificação de um conjunto de categorias preliminares sobre o fenômeno comum a todos eles, a comunicação dos dados de pesquisa. Tais categorias serão apresentadas na próxima seção deste capítulo.

4.2. Fatores que exercem influência na comunicação dos dados de pesquisa, apresentados pela literatura da ciência da informação

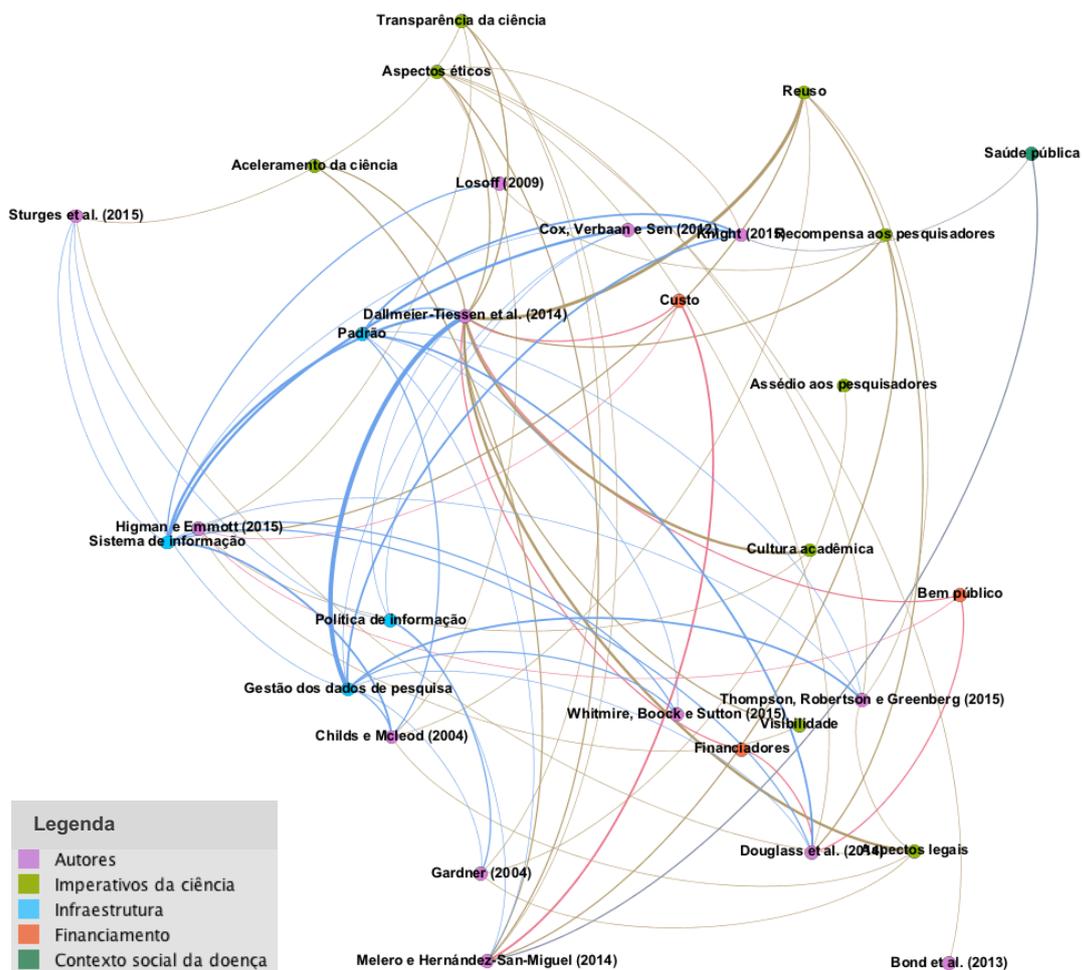
Neste estudo, a literatura científica no campo da ciência da informação foi utilizada como uma das fontes de informação a respeito dos fatores que influenciam a comunicação dos dados de pesquisa. Os elementos foram extraídos dos artigos por terem sido citados como pontos positivos, pontos negativos ou aspectos influenciadores nos processos relacionados à comunicação dos dados de pesquisa. Em alguns casos, a relação entre os elementos e seus efeitos não estava explicitamente descrita nos artigos, mas foi possível inferi-la por meio da interpretação do conteúdo.

Os elementos identificados deram origem às categorias preliminares sobre o tema investigado. O resultado foi obtido por meio da execução de três passos. O primeiro consistiu na rotularização livre dos elementos. No segundo passo, todos os elementos foram traduzidos para termos semelhantes denominados categorias

preliminares. Por fim, realizou-se o agrupamento das categorias em dimensões de análise (Apêndice A).

A sistematização dos elementos permitiu a identificação de 17 categorias preliminares. O conjunto das categorias foi organizado segundo as dimensões de análise do fenômeno descritas no modelo conceitual da pesquisa. A apresentação da relação foi elaborada a partir do software Gephi⁴⁵, para elaboração de uma rede entre os autores, as categorias apresentadas e suas dimensões (Figura 36).

Figura 36: Categorias preliminares e dimensões de análise.



Fonte: Elaboração própria a partir do software Gephi.

⁴⁵ Fonte: <https://gephi.org/>

A figura gerada é resultante de uma metodologia de análise de redes, conforme explicou Jacomy et al. (2014). Tal rede representa a relação dos autores e das categorias preliminares citadas por eles. As categorias já estão classificadas em uma das quatro dimensões de análise, segundo a indicação da legenda. Cada artigo, representado por suas autorias, e cada categoria representam um nó na rede. A relação entre eles é expressa pelas linhas, dotadas de força. Portanto, nessa representação, quanto mais grossa a linha de ligação entre os nós, mais relevante é a relação evidenciada.

A representação proposta permite confirmar duas evidências consideradas pela pesquisa. A primeira delas é a centralidade da discussão de Dallmeier-Tiessen et al. (2014) para a composição das categorias preliminares deste estudo. A segunda é a confirmação das dimensões de análise propostas no modelo conceitual da pesquisa. Tais dimensões foram propostas de maneira preliminar com o objetivo de estruturar os principais temas acerca da comunicação dos dados de pesquisa. Após a análise sistematizada da literatura, concluiu-se que as dimensões são adequadas para o estudo, da forma como foi concebido. A partir da dimensão de análise, serão discutidas todas as categorias preliminares e suas propriedades.

4.2.1. Categorias preliminares acerca dos imperativos da ciência

O fenômeno investigado neste estudo relaciona-se com as iniciativas da ciência aberta. Os princípios que sustentam tal proposta estão fundamentados nas mesmas ideias que foram base para as revoluções científicas e que caracterizam a ciência moderna, conforme explicou Boulton (2013). Na análise resultante da revisão sistematizada da literatura, foram identificadas nove categorias que se relacionam conceitualmente com as prerrogativas da ciência moderna, são elas: aceleração da ciência, reuso dos dados de pesquisa, transparência da ciência, assédio aos pesquisadores, visibilidade, recompensa aos pesquisadores, cultura acadêmica, aspectos legais e aspectos éticos.

4.2.1.1. Aceleração da ciência

A construção histórica do conhecimento é uma das características da ciência moderna destacada pelo filósofo Comte-Sponville (2011). Ela pressupõe que um novo

conhecimento baseia-se em conhecimentos já registrados e comunicados. Com isso, a comunicação dos conhecimentos é entendida como um elemento que acelera o desenvolvimento da ciência. A noção da comunicação dos dados de pesquisa como um fator que promove o desenvolvimento científico com maior rapidez foi discutido por Melero; Hernández-San-Miguel (2014) e Dallmeier-Tiessen et al. (2014) (Quadro 17).

Quadro 17: Aceleração da ciência.

Autores	Codificação dos dados
Melero e Hernández-San-Miguel (2014)	Maximização de conhecimentos
	Maximização do uso dos dados
Dallmeier-Tiessen et al. (2014)	Redução do tempo e custo para produção de novos resultados

Fonte: Elaboração própria.

De acordo com os autores, iniciativas que favorecem a ampla divulgação dos dados de pesquisa promovem a maximização do uso dos dados para a produção de novos conhecimentos de forma otimizada e acelerada.

4.2.1.2. Reuso dos dados de pesquisa

O aceleração da ciência é possibilitado por meio do reuso dos dados, aspecto defendido por cinco artigos do conjunto analisado. No entanto, além dos benefícios relacionados ao favorecimento da acumulação de conhecimentos e da construção histórica da ciência, os autores destacaram aspectos que inviabilizam ou dificultam o processo de comunicação dos dados de pesquisa, os quais estão diretamente relacionados ao reuso dos dados.

Dallmeier-Tiessen et al. (2014), Thompson, Robertson e Greenberg (2015) e Douglass et al. (2014) enfatizaram os custos relacionados aos processos para a comunicação de dados. De acordo com os autores, a preparação dos dados para sua publicação demanda mais tempo e mais esforços na produção dos dados, visto que os dados de pesquisa que são divulgados precisam estar estruturados em alguns padrões e com maior nível de detalhamento do que a pesquisa que os produziu demandaria. Além do esforço adicional para a preparação dos dados, Douglass et al. (2014) e Higman e Emmott (2015) pontuam que, em muitos casos, os dados produzidos por uma pesquisa não terão aplicabilidade para outros contextos.

Além dos questionamentos quanto à utilidade dos dados para outras pesquisas, foram destacados receios dos pesquisadores com o processo. Childs e Mcleod (2004) consideraram que os pesquisadores se apresentam temerosos com respeito aos usos

que podem ser feitos a partir de seus dados. Apesar de os autores não especificarem a origem do temor dos pesquisadores, Dallmeier-Tiessen et al. (2014) apresentam pistas sobre a questão. Os autores afirmaram que o receio dos pesquisadores com relação aos possíveis reusos de seus dados relaciona-se com duas possibilidades. Uma delas refere-se à desqualificação dos seus dados e, por consequência, dos resultados obtidos em sua pesquisa. A outra diz respeito às possíveis divergências de análises resultantes dos dados. Com isso, os dados comunicados abertamente seriam utilizados para desqualificar ou evidenciar contradições dos resultados apresentados pelos pesquisadores que produziram os dados (Quadro 18).

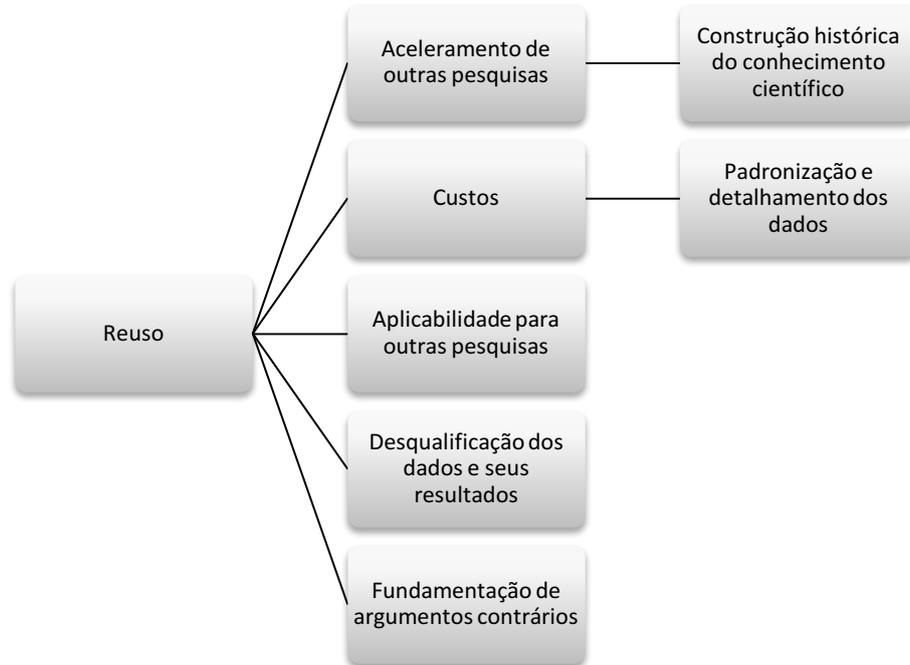
Quadro 18: Reuso dos dados.

Autores	Codificação dos dados
Childs e Mcleod (2004)	Propósito do reuso dos dados
Thompson, Robertson e Greenberg (2015)	Tempo para tornar os dados utilizáveis por outras pesquisas
Dallmeier-Tiessen et al. (2014)	Reutilização de dados
	Geração de novos resultados
	Esforço e tempo para preparar os dados para o reuso
	Desqualificação dos dados (receio dos pesquisadores)
	Divergência de análises (receio dos pesquisadores)
Higman e Emmott (2015)	Reuso dos dados
	Aplicabilidade do reuso
Douglass et al. (2014)	Tempo para tornar os dados reutilizáveis
	Aplicabilidade do reuso dos dados

Fonte: Elaboração própria.

O reuso dos dados de pesquisa geram aspectos (ou são gerados por eles) que podem ser entendidos ora como promotores, ora como inibidores da comunicação aberta dos dados de pesquisa. Sua versatilidade decorre da complexidade das questões envolvidas com a categoria. Portanto, suas vertentes foram explicitadas na Figura 37.

Figura 37: Desdobramentos da categoria “reuso dos dados”.



Fonte: Elaboração própria.

Em todos os casos, os argumentos estão relacionados a uma das características essenciais da ciência moderna, conforme fora explicitado por Comte-Sponville (2011). Tal característica diz respeito à refutabilidade do conhecimento científico, que demanda transparência nos processos de produção do conhecimento científico.

4.2.1.3. Transparência da ciência

A refutabilidade é outra característica da ciência moderna, amplamente discutida por teóricos como Karl Popper (2013) e Thomas Kuhn (2005). Cinco artigos do conjunto analisado argumentaram que a comunicação dos dados de pesquisa favorece, em alguma medida, o exercício dessa característica. Dallmeier-Tiessen et al. (2014) defenderam que a divulgação dos dados é importante para a validação dos resultados apresentados em um artigo ou outra forma de comunicação científica. Além disso, os autores destacaram que esses dados são úteis para testar as teorias propostas no futuro e embasar novas teorias. Portanto, os autores defendem que a comunicação dos dados de pesquisa é necessária para conferir integridade à ciência, uma vez que erros são mais facilmente detectados e a ciência é produzida com maior transparência (Quadro 19).

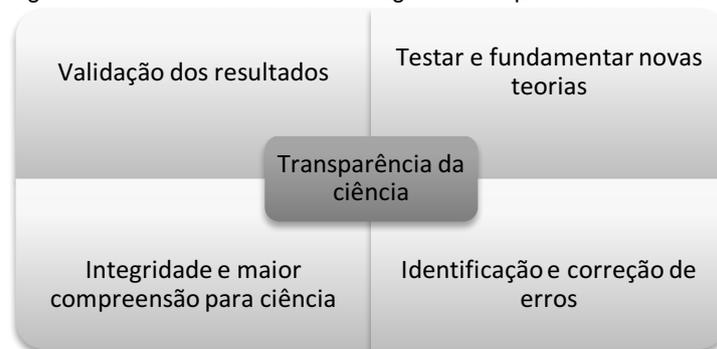
Quadro 19: Transparência da ciência.

Autores	Codificação dos dados
Sturges et al. (2015)	Transparência da ciência
Higman e Emmott (2015)	Transparência da ciência
Melero e Hernández-San-Miguel (2014)	Qualidade da ciência
Dallmeier-Tiessen et al. (2014)	Validação dos resultados
	Teste de teorias
	Integridade da ciência

Fonte: Elaboração própria.

O argumento do favorecimento da transparência da ciência por meio da comunicação dos dados de pesquisa também foi identificado nas discussões de Sturges et al. (2015) e Higman e Emmott (2015). Nesse sentido, Melero; Hernández-San-Miguel (2014) argumentam que o processo tem condições de proporcionar mais qualidade para a ciência produzida, uma vez que a possibilidade de diálogo com outras pesquisas é ampliada, bem como a capacidade de identificar erros e corrigi-los. Por fim, as considerações de Gardner (2004) sugerem que a ampla divulgação dos dados permite maior compreensão da pesquisa, portanto confere maior transparência (Figura 38).

Figura 38: Desdobramentos da categoria “transparência da ciência”.



Fonte: Elaboração própria.

Portanto, a categoria preliminar que apresenta o argumento de conferência de transparência para a ciência a partir da comunicação dos dados de pesquisa considera outros quatro benefícios correlatos. São eles: a validação dos resultados apresentados, a produção de novas teorias, a identificação de erros e a promoção de integridade para a ciência.

4.2.1.4. Assédio aos pesquisadores

No limiar da discussão entre a promoção da transparência da produção do conhecimento científico e os receios dos pesquisadores sobre os propósitos do reuso de seus dados, equilibra-se o aspecto apresentado por Gardner (2004) (Quadro 20).

Quadro 20: Assédio aos pesquisadores.

Autores	Codificação dos dados
Gardner (2004)	Assédio aos pesquisadores

Fonte: Elaboração própria.

De acordo com Gardner (2004), a legislação estadunidense que determina que todos os pesquisadores devem divulgar seus dados de pesquisa sempre que for solicitado é um fator que atrasa as pesquisas por deixar os pesquisadores a mercê dessas demandas. Além disso, o autor denunciou que as solicitações têm representado um instrumento para assediar os pesquisadores que produzem conhecimento que contraria o interesse de grandes corporações comerciais, tais como a indústria farmacêutica.

4.2.1.5. Visibilidade

A ampliação da visibilidade da pesquisa é um fator que favorece o interesse por parte dos pesquisadores e suas instituições para promoverem a comunicação dos dados de pesquisa, conforme ressaltaram Dallmeier-Tiessen et al. (2014) e Higman e Emmott (2015). Para os autores, a ampliação da divulgação dos dados aumenta também a possibilidade de reconhecimento do trabalho e de atribuição da autoria àqueles resultados (Quadro 21).

Quadro 21: Visibilidade.

Autores	Codificação dos dados
Dallmeier-Tiessen et al. (2014)	Reconhecimento da produção
	Ampliação da visibilidade (citação)
Higman e Emmott (2015)	Visibilidade da pesquisa

Fonte: Elaboração própria.

Uma das formas de expressão da visibilidade é aferida por meio da taxa de citação do trabalho (Figura 39).

Figura 39: Desdobramentos da categoria “visibilidade”.



Fonte: Elaboração própria.

No contexto da ciência aberta, argumenta-se que, quanto mais abertamente o trabalho for comunicado, maior será seu alcance e tão maior será sua taxa de citação. Até o momento, não foi localizado nenhum estudo que tenha confirmado ou refutado essa relação para o contexto dos dados de pesquisa. Entretanto, o argumento é recorrente nas discussões acerca do acesso aberto às publicações científicas.

Apesar da recorrência do argumento, o tema ainda se mostra controverso. Droscher (2012) investigou a relação entre abertura na divulgação e aumento na taxa de citação de artigos de periódicos. Entre os resultados da pesquisa, foi observado que acesso não é condição suficiente para que haja citação. Dessa forma, com base em seus resultados, a autora não confirmou que o acesso aberto tenha como consequência direta o aumento do impacto das publicações. Apesar disso, a autora pondera que o acesso aberto aumenta drasticamente o número de usuários potenciais e que o crescimento de uso pode aumentar o impacto da pesquisa em sua área do conhecimento.

Embora a relação entre acesso aberto e aumento na taxa de citação não seja unânime, acredita-se que a ampliação do acesso aos produtos da pesquisa resulta na ampliação da visibilidade da pesquisa. Embora a comparação seja plausível para a natureza da discussão, a comunicação dos dados de pesquisa apresenta

especificidades com relação às publicações científicas, tais como artigos de periódicos, especialmente no que diz respeito às recompensas recebidas pelos pesquisadores.

4.2.1.6. Recompensa aos pesquisadores

Diferentemente da comunicação via artigos de periódicos, os dados de pesquisa não pontuam como produção acadêmica para os pesquisadores. Segundo Dallmeier-Tiessen et al. (2014), os pesquisadores alegaram que o compartilhamento dos dados de pesquisa não gera status nem promoção em suas carreiras, portanto não compensava o esforço. O tema foi pontuado pelos artigos de Melero, Hernández-San-Miguel (2014), Dallmeier-Tiessen et al. (2014) e Losoff (2009). De forma geral, os autores destacaram que o baixo valor acadêmico que é atribuído à divulgação dos dados somado à falta de incentivos para a prática são aspectos que inibem o processo (Quadro 22).

Quadro 22: Recompensas aos pesquisadores.

Autores	Codificação dos dados
Losoff (2009)	Recompensa aos autores pela divulgação dos dados
Dallmeier-Tiessen et al. (2014)	Status e promoção da carreira de pesquisador
	Valor acadêmico
Melero e Hernández-San-Miguel (2014)	Incentivo
	Interesse dos pesquisadores

Fonte: Elaboração própria.

É necessário ressaltar que a falta de valor acadêmico refere-se apenas ao processo de divulgação, que compõe um dos três macroprocessos que envolve a noção de comunicação de dados de pesquisa adotada neste estudo. Portanto, admite-se que a produção e o uso dos dados de pesquisa são processos valorizados e até mesmo requeridos para a promoção da carreira acadêmica dos pesquisadores.

4.2.1.7. Cultura acadêmica

Outro aspecto relacionado diretamente com os imperativos da ciência moderna é a cultura acadêmica em voga. O artigo de Dallmeier-Tiessen et al. (2014) foi o único a problematizar a temática explicitamente. Os autores consideraram dois aspectos como influenciadores nas práticas da comunicação dos dados de pesquisa: a cultura predominante na área e a idade dos pesquisadores (Quadro 23).

Quadro 23: Cultura acadêmica.

Autores	Codificação dos dados
Dallmeier-Tiessen et al. (2014)	Cultura acadêmica
	Idade dos pesquisadores
	Interação com outras disciplinas
	Interdisciplinaridade dos dados

Fonte: Elaboração própria.

Segundo Dallmeier-Tiessen et al. (2014), a cultura é a norma de comportamento manifestada no comportamento dos pesquisadores. A depender da área do conhecimento, existirão normas de comportamento mais ou menos propícias para a ampla comunicação dos dados de pesquisa. No entanto, os autores ponderam que, em muitos casos, são os interesses comerciais envolvidos na pesquisa que determinarão o comportamento dos pesquisadores em relação à divulgação dos dados (Figura 40).

Figura 40: Desdobramentos da categoria “cultura acadêmica”.



Fonte: Elaboração própria.

Somadas às normas de comportamento e aos interesses comerciais, os autores enfatizaram que o uso de tecnologias de informação e comunicação pode promover práticas mais colaborativas para a ciência, facilitando a interdisciplinaridade entre as pesquisas e a abertura na divulgação dos dados de pesquisa. De acordo com os autores, o uso das tecnologias com tal finalidade é mais recorrente entre pesquisadores jovens. Portanto, sob o aspecto da cultura acadêmica ainda foi adicionado mais um fator

que se refere à idade do pesquisador. A partir dessa perspectiva, pesquisadores mais jovens seriam mais propícios a comunicar abertamente seus dados de pesquisa, enquanto os pesquisadores mais velhos seriam mais resistentes a tais práticas.

4.2.1.8. Aspectos legais

Os aspectos legais envolvem as aplicações jurídicas nos três macroprocessos da comunicação dos dados de pesquisa. As legislações envolvidas nas discussões são diversas e por vezes conflituosas, conforme pontuou Dallmeier-Tiessen et al. (2014). De acordo com os autores, as divergências legais são decorrentes das dissonâncias entre as políticas de diferentes agências de fomento, países e regiões.

Childs e Mcleod (2004) e Dallmeier-Tiessen et al. (2014) ressaltaram alguns aspectos acerca das implicações dos direitos de propriedade intelectual na comunicação dos dados de pesquisa. Segundo Childs e Mcleod (2004), tais direitos podem regulamentar as práticas de armazenamento, acesso e reuso dos dados de pesquisa. Os autores não exploraram as nuances da relação, apenas as apontaram. No mesmo sentido, Sturges et al. (2015), Gardner (2004) e Thompson, Robertson e Greenberg (2015) apresentaram a temática como uma das barreiras para a ampla comunicação dos dados de pesquisa (Quadro 24).

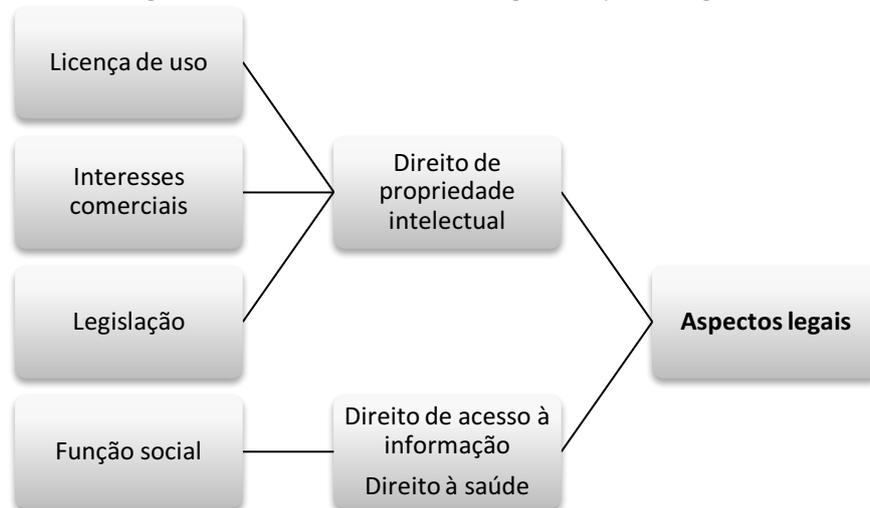
Quadro 24: Aspectos legais.

Autores	Codificação dos dados
Childs e Mcleod (2004)	Implicações legais
Sturges et al. (2015)	Licenças de uso
Gardner (2004)	Clareza das legislações
Dallmeier-Tiessen et al. (2014)	Direito de acesso à informação
	Interesses comerciais
	Direito de propriedade intelectual
	Legislação
Thompson, Robertson e Greenberg (2015)	Proteção dos direitos de propriedade intelectual

Fonte: Elaboração própria

O tema mostra-se como fortemente interveniente no processo investigado, uma vez que a legislação e outros aspectos legais têm força jurídica para impedir completamente a divulgação dos dados ou para torná-la compulsória (Figura 41).

Figura 41: Desdobramentos da categoria “aspectos legais”.



Fonte: Elaboração própria.

Em razão da relevância do tema para o contexto do estudo, o tópico será analisado extrapolando os limites das discussões dos autores que o citaram. Para tanto, serão apresentadas as considerações dos seguintes autores sobre o tema: Diniz (2005), Adolfo (2006), Branco (2007), Pinto (2009) e Mizukami (2007).

Direitos de propriedade intelectual

Os direitos de propriedade intelectual versam sobre os produtos da criação intelectual humana. Diniz (2005) conceitua-o como um direito imaterial, pois se refere à regulação de propriedade imaterial decorrente do trabalho intelectual, tais como obras literárias, artísticas, científicas e tecnológicas. Os processos de regulação jurídica para essas obras tiveram início em dois momentos históricos – a Revolução Francesa (1789-1799) e a Revolução Industrial (1760-1840). Ambos propiciaram a noção da criação intelectual, ou seja, um produto imaterial. De acordo com Adolfo (2006), os princípios éticos, o sistema jurídico e o ideal de Estado democrático advindos da Revolução Francesa foram as condições objetivas para constituição da noção de propriedade imaterial e intelectual. Já a Revolução Industrial institucionalizou as relações jurídicas no âmbito da propriedade intelectual, especialmente no contexto industrial, a partir de 1474, com a formalização do sistema de patentes, conforme pontuou Branco (2007).

Os direitos de propriedade intelectual são segmentados em dois domínios. O primeiro deles diz respeito aos direitos de propriedade industrial que atuam no âmbito dos bens industriais, tais como marcas, patentes e modelos. No Brasil, tais direitos são regulados pela Lei nº 9.279, publicada em 1996. No outro domínio, estão os direitos autorais, que abarcam os direitos de autor, direitos conexos⁴⁶ e regulamentação jurídica para os programas computacionais. Diferentemente dos direitos de propriedade industrial, os direitos autorais são objeto de estudo do direito civil. No Brasil, os direitos autorais são regulados pelas leis nº 9.610 e nº 9.609, ambas publicadas em 1998.

Além das legislações específicas já citadas, a própria Constituição Federal de 1988 garante os direitos de propriedade intelectual. No artigo 5º, que trata dos direitos fundamentais, é assegurado que:

XXVII - aos autores pertence o direito exclusivo de utilização, publicação ou reprodução de suas obras, transmissível aos herdeiros pelo tempo que a lei fixar;

XXVIII - são assegurados, nos termos da lei:

a) a proteção às participações individuais em obras coletivas e à reprodução da imagem e voz humanas, inclusive nas atividades desportivas;

b) o direito de fiscalização do aproveitamento econômico das obras que criarem ou de que participarem aos criadores, aos intérpretes e às respectivas representações sindicais e associativas;

XXIX - a lei assegurará aos autores de inventos industriais privilégio temporário para sua utilização, bem como proteção às criações industriais, à propriedade das marcas, aos nomes de empresas e a outros signos distintivos, tendo em vista o interesse social e o desenvolvimento tecnológico e econômico do País (BRASIL, 1988).

As duas vertentes do direito de propriedade intelectual apresentadas aqui são aplicáveis ao contexto dos dados de pesquisa, especialmente àquele que trata do desenvolvimento de fármacos. Isso porque os dados podem ser protegidos ou terem seu acesso limitado por causa de interesses comerciais e industriais ou ainda em decorrência dos interesses dos autores e titulares de direitos sobre os resultados das pesquisas.

Apesar da aparente consolidação das garantias à propriedade intelectual, o tema é controverso e polêmico, visto que, na ambivalência de tais direitos, situa-se a defesa do amplo acesso à informação. Portanto, faz-se necessário explorar outra abordagem concernente aos aspectos legais que envolvem a comunicação dos dados de pesquisa.

⁴⁶ Segundo Jungmann (2010), os direitos conexos referem-se à proteção para artistas intérpretes ou executantes, produtores fonográficos e empresas de radiodifusão, em decorrência de interpretação, execução, gravação ou veiculação das suas interpretações e execuções.

Direito ao acesso à informação

O direito à informação, tal como o direito a propriedade, também constitui um direito fundamental do cidadão. Ambos estão garantidos no artigo 5º da Constituição Federal de 1988 (BRASIL, 1988) e em acordos internacionais específicos do tema. A abordagem amplamente discutida sobre os direitos de acesso à informação abrange o âmbito público e recentemente, o universo circunscrito da propriedade intelectual (PINTO, 2009). A ampliação da abordagem fundamenta-se na perspectiva de que a propriedade intelectual deve cumprir uma função social, tal como os outros tipos de propriedade. Com isso, Pinto (2009) afirma que, recorrer ao direito de acesso à informação para problematizar o direito de propriedade intelectual, não é negá-lo, e sim torná-lo um direito coletivo.

A visão de coletividade para a propriedade intelectual advém da percepção da produção intelectual como um processo coletivo. Segundo Pinto (2009), ela parte da “concepção de que o conhecimento produzido pelo homem é resultado da contribuição de seus antepassados e que sua apropriação individual denega a sua origem coletiva”.

No mesmo sentido de problematizar a relação, Mizukami (2007) sugere que a propriedade intelectual visa garantir um direito de exploração comercial da obra, que nem sempre beneficia o produtor ou sua obra. Portanto, o autor defende que os direitos devem ser considerados dentro de uma dimensão social que seja capaz de balanceá-los.

A partir da discussão aqui apresentada, enfatiza-se que o direito de propriedade intelectual, embora seja reconhecido, não é um direito absoluto. Assim como as demais propriedades, ele deve exercer uma função social, a qual deve garantir o exercício de outros direitos, como o direito à informação, à educação e à saúde (PINTO, 2009).

4.2.1.9. Aspectos éticos

Os aspectos éticos foram recorrentes nas discussões dos artigos analisados. Eles estavam declaradamente presentes em 8 dos 13 artigos analisados. No entanto, observou-se que tais aspectos estão centrados nas preocupações com a privacidade

dos participantes (Childs e Mcleod, 2004; Dallmeier-Tiessen et al., 2014; Melero e Hernández-San-Miguel, 2014; Thompson, Robertson e Greenberg, 2015). Duas subcategorias foram identificadas a partir da discussão: a anonimização dos dados e a obtenção da permissão dos participantes. Ambas foram apresentadas como alternativas para permitir a comunicação aberta dos dados e manter os direitos à privacidade dos participantes (Quadro 25).

Quadro 25: Aspectos éticos.

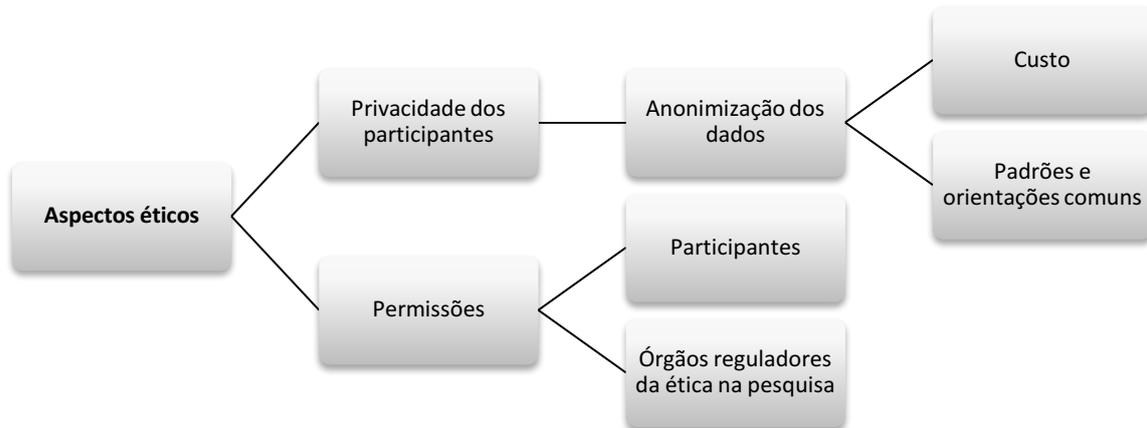
Autores	Codificação dos dados
Dallmeier-Tiessen et al. (2014)	Privacidade dos participantes
	Custo e padrão para anonimizar os dados
Douglass et al. (2014)	Direitos para disponibilizar os dados
Melero e Hernández-San-Miguel (2014)	Privacidade dos participantes
	Anonimato dos dados
Bond et al. (2013)	Permissão para o uso dos dados
Childs e Mcleod (2004)	Confidencialidade dos dados
Thompson, Robertson e Greenberg (2015)	Privacidade dos participantes
Knight (2015)	Permissão dos participantes da pesquisa

Fonte: Elaboração própria.

A anonimização dos dados como estratégia foi apontada por Dallmeier-Tiessen et al. (2014) e Melero e Hernández-San-Miguel (2014). No entanto, Dallmeier-Tiessen et al. (2014) levantaram preocupações acerca dos custos envolvidos, que podem ser proibitivos para projetos de pesquisas com recursos reduzidos. Somados aos custos, apresenta-se ainda como barreira a ausência de padrões e de orientação que sejam amplamente aceitos para tal finalidade.

Outra estratégia identificada nas discussões sugeriu que a falta de autorizações para a divulgação e o reuso dos dados dos participantes é um impeditivo para a comunicação aberta dos dados de pesquisa (Douglass et al., 2014; Bond et al., 2013 e Knight, 2015). Segundo essa proposta, as permissões deveriam ser obtidas a partir dos órgãos reguladores de ética na pesquisa e de autorizações dos próprios participantes (Figura 42).

Figura 42: Desdobramentos da categoria “aspectos éticos”.



Fonte: Elaboração própria.

De forma geral, a necessidade de manter a privacidade dos participantes seria um impedimento para a ampla comunicação dos dados de pesquisa. No entanto, na discussão ampliada elaborada a partir dos artigos selecionados, surgiu um novo argumento para a temática, na qual a comunicação aberta dos dados de pesquisa é um cumprimento de exigências éticas para pesquisas no campo das ciências da saúde. Pelo fato de tal perspectiva estar relacionada exclusivamente com uma área do conhecimento, o qual é objeto dessa investigação, o tema será explorado na próxima seção do capítulo.

4.2.2. Categorias preliminares acerca do contexto social da doença

As pesquisas nas ciências da saúde partem das mesmas prerrogativas da ciência moderna, caracterizadas no tópico anterior. No entanto, a área possui algumas especificidades que são de especial interesse para o estudo. Suas singularidades favorecem a comunicação aberta dos dados de pesquisa, conforme sugeriu Knight (2015), Melero e Hernández-San-Miguel (2014) e Bond et al. (2013). Adicionalmente, os argumentos em defesa da ciência aberta na área do conhecimento ganham mais legitimidade em alguns temas de amplo interesse social, como é o caso da saúde

pública. Portanto, as categorias preliminares para fundamentar a discussão sobre o contexto social da doença trataram diretamente do prisma da saúde pública.

4.2.2.1. Saúde pública

Dois artigos ressaltaram o contexto das pesquisas em saúde pública como impulsionador da comunicação aberta dos dados de pesquisa. O mais recente foi publicado por Knight (2015). Embora a relação entre a saúde pública e a comunicação dos dados de pesquisa não seja explorada diretamente, ela é evidenciada pela experiência relatada. A autora afirma que o serviço de gestão dos dados de pesquisa da universidade é ofertado para todos os institutos. No entanto, aquele que trata de medicina tropical é protagonista da discussão e ampliação do serviço.

A relação também é estabelecida de forma indireta no artigo de Melero e Hernández-San-Miguel (2014). A partir dos resultados obtidos em dois diretórios internacionais de políticas dessa natureza, os autores concluíram que as áreas das ciências da saúde apresentam mais políticas de informação para o acesso aberto do que as outras áreas do conhecimento (Quadro 26).

Quadro 26: Saúde pública.

Autores	Codificação dos dados
Knight (2015)	Contexto das pesquisas em saúde pública
Melero e Hernández-San-Miguel (2014)	Ciências da saúde
	Progresso da saúde pública

Fonte: Elaboração própria.

Segundo Melero e Hernández-San-Miguel (2014), a influência da política da agência Wellcome Trust (2011) é um dos elementos que configuram tal cenário. Portanto, faz-se necessário destacar os argumentos apresentados pela iniciativa e pelas discussões que a circundam. Segundo a política da Wellcome Trust (2011), a ampla comunicação dos dados de pesquisa acelera o progresso da saúde pública, melhorando assim a qualidade de vida das populações. O efeito, por si só, já é de grande interesse social e cumpre uma importante função social. No mesmo ano de publicação da política, Walport e Brest (2011) subscreveram-na e acrescentaram que a comunicação aberta dos dados de pesquisa maximiza o uso dos dados coletados e, por consequência, a criação de conhecimento a partir deles. De acordo com os atores, a maximização do uso dos dados, especialmente daqueles coletados das populações, é um princípio ético para as pesquisas na área das ciências da saúde, uma vez que as populações não

devem ser submetidas novamente à coleta de dados que já poderiam estar disponíveis (Figura 43).

Figura 43: Desdobramentos da categoria “saúde pública”.



Fonte: Elaboração própria.

Portanto, a categoria preliminar acerca da saúde pública relaciona-se com três elementos de qualificação. O primeiro deles abarca as políticas de informação para a promoção da comunicação dos dados de pesquisa. Conforme a discussão levantada pela o estudo, as ciências da saúde são percussoras e exitosas na promoção de tais políticas. O segundo elemento diz respeito a função social da produção de conhecimento científico, que especialmente em temáticas relacionadas a saúde pública traz como compromisso o melhoramento da qualidade de vida das populações atingidas pelo estudo. Por fim, o terceiro elemento trata sobre a maximização do uso dos dados como um princípio ético dos estudos relacionados a saúde pública. Isso porque as populações acometidas pelo tema da saúde pública devem ser preservadas e a coleta dos dados deve ser minimizado ao máximo possível.

4.2.3. Categorias preliminares acerca do financiamento

A experiência da agência Wellcome Trust (2011) e sua influêncica para o contexto das ciências da saúde ilustram a relevância das agências de fomento à pesquisa para

a condução da abertura da comunicação dos dados. Nos casos em que o financiamento é de natureza pública, ressalta-se a noção dos resultados como um bem público, ou seja, pertencente a toda sociedade. Tal perspectiva extrapola os pressupostos da ciência moderna de publicização dos resultados e busca se apresentar como um direito social.

4.2.3.1. Instituições financiadoras de pesquisa

As agências financiadoras de pesquisa, bem como as instituições de ensino superior e pesquisa, são consideradas agentes estratégicos para garantir a comunicação dos dados de pesquisa, conforme discutiu Dallmeier-Tiessen et al. (2014) e Douglass et al. (2014) (Quadro 27).

Quadro 27: Instituição financiadora.

Autores	Codificação dos dados
Dallmeier-Tiessen et al. (2014)	Requisito para financiamento de pesquisa
	Crescimento de financiamento
Douglass et al. (2014)	Financiamento
	Políticas mandatórias das financiadoras

Fonte: Elaboração própria.

De forma geral, os autores defenderam que as instituições têm autoridade para requisitar parâmetros para coleta, divulgação e uso dos dados das pesquisas que financiam (Figura 44).

Figura 44: Desdobramentos da categoria “instituição financiadora”.



Fonte: Elaboração própria.

Douglass et al. (2014) indicam as políticas de informação mandatórias como instrumentos estratégicos para estabelecer os requisitos do financiamento à promoção da comunicação dos dados de pesquisa de forma aberta. Segundo Dallmeier-Tiessen et al. (2014), isso permitira que as instituições financiadoras de pesquisa otimizassem os recursos investidos. A perspectiva otimista defendida por Dallmeier-Tiessen et al. (2014) baseia-se na expectativa de que a comunicação dos dados de pesquisa tem potencial para promover uma economia nos recursos investidos na pesquisa em longo prazo.

4.2.3.2. Bem público

Além dos elementos discutidos na categoria “aspectos legais sobre o direito de acesso à informação”, alguns autores pontuaram que pesquisas financiadas com recursos públicos geram, por excelência, bens públicos. Uma vez entendidos os dados resultantes dessas pesquisas como bens públicos, eles deveriam estar amplamente disponíveis para quem deles necessitassem, para tanto deveriam ser objetos de uma comunicação aberta. O argumento pode ser observado, de forma superficial, no artigo

de Dallmeier-Tiessen et al. (2014) e Higman e Emmott (2015) e, de forma mais aprofundada, no artigo de Douglass et al. (2014) (Quadro 28).

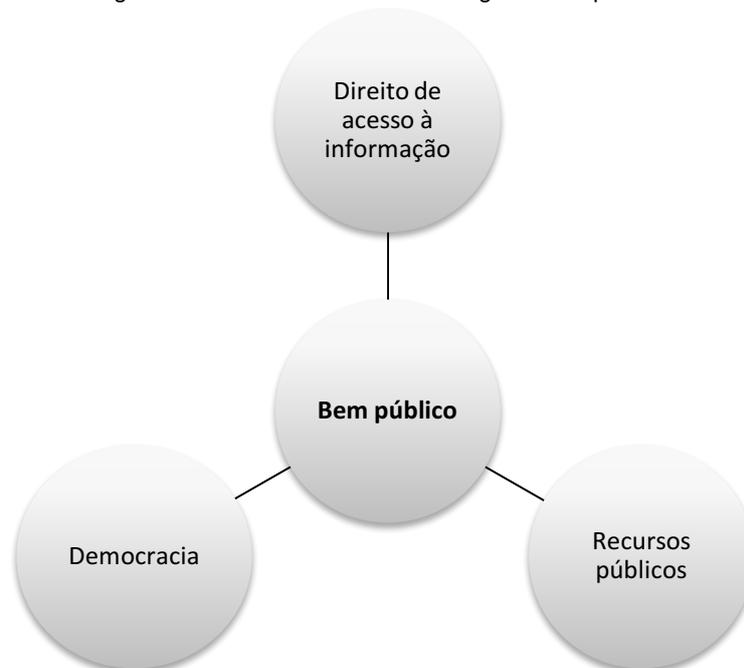
Quadro 28: Bem público.

Autores	Codificação dos dados
Dallmeier-Tiessen et al. (2014)	Benefícios sociais
	Entendimento público sobre as pesquisas
Higman e Emmott (2015)	Recursos públicos e bens públicos
Douglass et al. (2014)	Democracia
	Interesse em tornar os dados públicos

Fonte: Elaboração própria.

Douglass et al. (2014) discutiram o tema a partir das políticas de comunicação de dados do governo norte-americano. Os autores alegaram que especialmente os dados de pesquisas e de agências governamentais obtidos com recursos públicos devem ser livremente disponibilizados para o público. A defesa da alegação, entre outros aspectos, baseou-se no pronunciamento do então presidente dos Estados Unidos, que afirmou que a prática seria um elemento que fortaleceria a democracia do país (Figura 45).

Figura 45: Desdobramentos da categoria bem público.



Fonte: Elaboração própria.

A noção do conhecimento científico como bem público, especialmente aquele produzido com recursos públicos acende a reivindicação de dois direitos fundamentais. Um deles é a democracia e o outro é o direito à informação.

4.2.3.3. Custo

A discussão acerca do custo da comunicação dos dados de pesquisa centrou-se nos aspectos relativos aos investimentos para infraestrutura necessária, que é objeto de financiamento ou demanda recursos para estruturá-la. Dallmeier-Tiessen et al. (2014) e Higman e Emmott (2015) apontaram que os custos para a manutenção da infraestrutura podem ser elevados, mas esse investimento é altamente relevante para os objetivos da comunicação. Portanto, os custos envolvidos nos processos e na infraestrutura para suportar a comunicação dos dados de pesquisa precisam ser levados em consideração no financiamento da pesquisa. Dallmeier-Tiessen et al. (2014) enfatizam que a requisição para a divulgação dos dados pode ser uma forma de controlar os recursos investidos. No entanto, os autores pontuam que os custos devem ser considerados também para longo prazo, de forma a garantir a preservação dos recursos.

Além das considerações acerca do custo para infraestrutura do sistema de informação, Melero e Hernández-San-Miguel (2014) e autores relacionados enfatizaram dois aspectos igualmente relevantes para a análise aqui proposta. Um deles relaciona-se com a economia dos recursos e o outro com os custos da coleta e análise dos dados (Quadro 29).

Quadro 29: Custo.

Autores	Codificação dos dados
Melero e Hernández-San-Miguel (2014)	Custo (tempo e dinheiro) para coleta e análise dos dados Aproveitamento de recursos
Higman e Emmott (2015)	Altos custos
Dallmeier-Tiessen et al. (2014)	Controle dos recursos Custo e manutenção da infraestrutura

Fonte: Elaboração própria.

Os dois aspectos relacionam-se com a iniciativa da agência Wellcome Trust (2011) que visou consolidar um compromisso para a comunicação aberta dos dados de pesquisa entre as instituições financiadoras de pesquisa no campo das ciências da saúde (Figura 46).

Figura 46: Desdobramentos da categoria “custo”.



Fonte: Elaboração própria.

Entre os principais argumentos utilizados pela agência para tal acordo, estava a perspectiva de que a ampla divulgação dos dados de pesquisa gera um melhor aproveitamento dos recursos investidos, portanto é um elemento promotor de economia dos financiamentos.

Outros argumentos apresentados pela agência sugerem que a prática de comunicação aberta dos dados de pesquisa gera o uso justo dos recursos da ciência. No entanto, uma das perspectivas apresentadas nessa tese sobre a temática afirma que alguns países possuem mais condições de arcar com os custos da análise dos dados do que outros. Portanto, o uso justo dos dados de alguma forma implicaria a destinação de recursos para a análise dos dados, de forma a permitir que todos os pesquisadores tenham as mesmas condições tanto de coleta como de análise dos dados.

4.2.4. Categorias preliminares acerca da infraestrutura

Um dos elementos centrais relacionados ao financiamento da pesquisa e da comunicação de dados diz respeito aos custos com a infraestrutura. Nessa dimensão, estão reunidas categorias que permitem a expressão e operacionalização do fenômeno

investigado. Em decorrência do universo da discussão – a ciência da informação –, as categorias mais presentes nos artigos analisados serão descritas a seguir.

4.2.4.1. Sistemas de informação

A análise aqui empregada permitiu identificar que os aspectos referentes aos sistemas de informação foi o fator mais citado na literatura selecionada. Sob a categoria “sistemas de informação”, foram classificados os temas referentes aos repositórios de dados de pesquisa, aos periódicos, aos sistemas de descoberta e segurança, à tecnologia e à interoperabilidade dos sistemas de divulgação dos dados de pesquisa (Figura 47).

Figura 47: Desdobramentos da categoria “sistema de informação”.



Fonte: Elaboração própria.

Compreende-se da análise que os repositórios de dados de pesquisa são alternativas eficientes para o depósito e armazenamento dos dados, capazes de absorver a diversidade do volume e os tipos de arquivos que os dados de pesquisa podem apresentar (Losoff, 2009; Sturges et al., 2015). No entanto, foi levantada a preocupação com a garantia da preservação e da segurança dos dados e a capacidade de integração desses sistemas com outros, tais como os periódicos científicos (Losoff,

2009), a tecnologia envolvida (Childs e MCleod, 2004) e os sistemas de busca (Dallmeier-Tiessen et al., 2014; Douglass et al., 2014).

Outros autores como Thompson, Robertson e Greenberg (2015) e Knight (2015) indicaram concordância em relação à necessidade de sistemas de informação destinados à operacionalização da comunicação dos dados de pesquisa. No entanto, os autores não envergaram nem para os repositórios de dados tão pouco para os periódicos. Os apontamentos dos autores foram apenas no intuito de evidenciar a necessidade e pontuar, de maneira geral, que os sistemas deveriam ser seguros e interoperáveis (Quadro 30).

Quadro 30: Sistemas de informação.

Autores	Codificação dos dados
Losoff (2009)	Periódicos para dados de pesquisa
Cox, Verbaan e Sen (2012)	Sistemas para a comunicação dos dados de pesquisa
Whitmire, Boock e Sutton (2015)	Armazenamento dos dados (volume e preservação)
	Central de armazenamento dos dados
Thompson, Robertson e Greenberg (2015)	Acesso (mídias e repositórios)
	Sistemas de descoberta dos dados
Dallmeier-Tiessen et al. (2014)	Falta de segurança
	Integração dos sistemas
	Relação dos dados com as publicações científicas
Douglass et al. (2014)	Motores de busca
	Local de depósito
Knight (2015)	Segurança
	Local de armazenamento
	Sistemas utilizados
	Acesso (velocidade)
Sturges et al. (2015)	Repositório de dados de pesquisa
Childs e Mcleod (2004)	Sistemas tecnológicos
	Infraestrutura tecnológica
	Agrupamento dos dados em datasets

Fonte: Elaboração própria.

Nenhuma característica a respeito dos repositórios ou periódicos foi explorada pela literatura analisada. Portanto, para o aspecto relacionado ao sistema de informação, será considerada a explanação realizada na revisão de literatura deste trabalho, que tratou dos repositórios de dados de pesquisa (pág. 46).

4.2.4.2. Padrões

A segunda categoria mais presente entre os artigos refere-se aos padrões necessários aos processos, procedimentos e sistemas que envolvem a comunicação dos dados de pesquisa (Quadro 31).

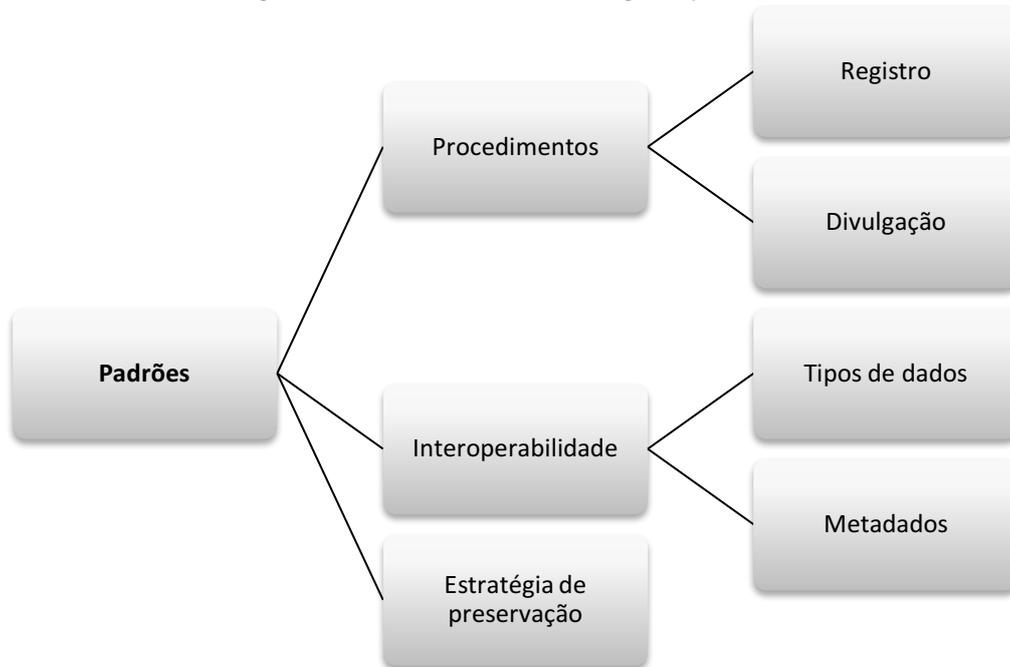
Quadro 31: Padrões.

Autores	Codificação dos dados
Childs e Mcleod (2004)	Padrão de processos e procedimentos
	Indicações de padrões e políticas
Whitmire, Boock e Sutton (2015)	Padrões de dados
Thompson, Robertson e Greenberg (2015)	Padrões de metadados
Dallmeier-Tiessen et al. (2014)	Interoperabilidade dos dados
	Padrões para a descrição dos dados
Douglass et al. (2014)	Padrões
	Metadados (padrões)
	Interoperabilidade
Knight (2015)	Padrão
	Interoperabilidade
	Formato dos dados
Melero e Hernández-San-Miguel (2014)	Padrões

Fonte: Elaboração própria.

Childs e Mcleod (2004) e Whitmire, Boock e Sutton (2015) discutiram duas naturezas de padronização, uma relacionada aos procedimentos e outra ligada ao formato dos dados. De acordo com o levantamento realizado por Childs e Mcleod (2004), a falta de padronização nos procedimentos relacionados com o registro e a divulgação dos é um grave problema para a comunicação dos dados na perspectiva de pesquisadores. Vale ressaltar que a investigação empenhada pelos autores apontou tal barreira a partir da perspectiva de pesquisadores. Com isso, os autores concluíram que a prestação de serviços de orientações aos pesquisadores, bem como a indicação de padrões são fundamentais para o sucesso do projeto analisado. Whitmire, Boock e Sutton (2015) não problematizaram a questão da importância dos padrões, no entanto sua análise fundamenta-se na classificação do tipo de dados de pesquisa por área do conhecimento e, portanto, pressupõe a lógica de padrão para a tipologia dos dados s (Figura 48).

Figura 48: Desdobramentos da categoria “padrões”,



Fonte: Elaboração própria.

Os outros autores que abordaram a temática dos padrões (Thompson, Robertson, Greenberg, 2015; Dallmeier-Tiessen et al., 2014; Douglass et al., 2014; Knight, 2015) focaram no aspecto dos metadados. Thompson, Robertson e Greenberg (2015) indicaram o uso do padrão de metadados como elemento necessário à adoção de estratégia de preservação dos dados de pesquisa. A mesma abordagem foi adotada por Dallmeier-Tiessen et al. (2014), que afirmaram que a falta de padrão para os metadados constitui uma barreira para a efetiva preservação dos dados de pesquisa. Douglass et al. (2014) e Knight (2015) consideraram que a incerteza sobre os padrões de metadados a serem adotados no processo de comunicação dos dados de pesquisa é impeditiva para o processo e reflete-se na capacidade de interoperabilidade dos dados.

4.2.4.3. Gestão dos dados de pesquisa

A discussão acerca da gestão dos dados de pesquisa foi desenvolvida por oito artigos analisados (Quadro 32). De forma geral, todos eles apontaram para a necessidade da existência de um serviço de informação ou ressaltaram sua importância

em contextos onde este já se faz presente, conforme relatado por Knight (2015) e Whitmire, Boock e Sutton (2015).

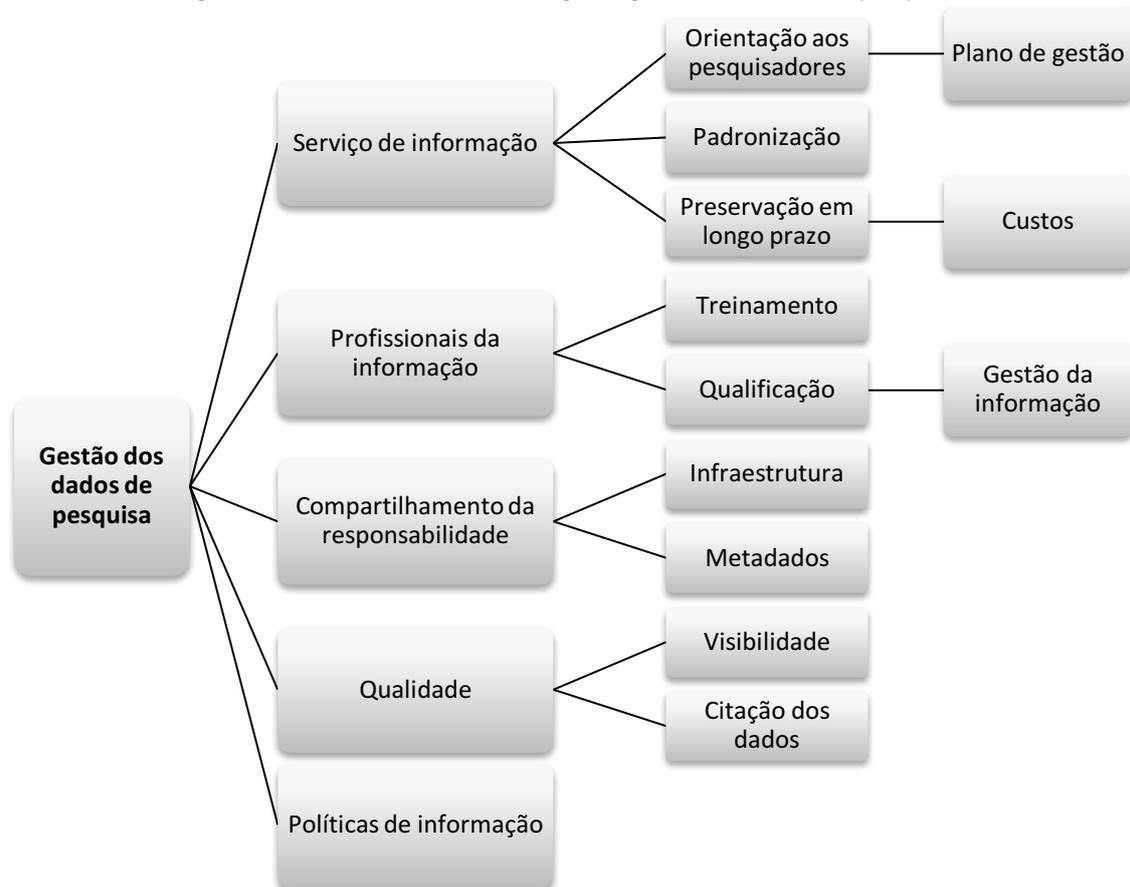
Quadro 32: Gestão dos dados de pesquisa.

Autores	Codificação dos dados
Cox, Verbaan e Sen (2012)	Gestão dos dados de pesquisa
Whitmire, Boock e Sutton (2015)	Responsabilidades com a gestão dos dados
	Serviço de gestão dos dados
Gardner (2004)	Responsabilidade compartilhada
Knight (2015)	Práticas de arquivamento (desconhecimento)
	Plano de gestão dos dados
	Serviço para gestão dos dados
Thompson, Robertson e Greenberg (2015)	Preservação dos dados
	Preservação (padrões e políticas)
	Responsáveis pela curadoria dos dados
Dallmeier-Tiessen et al. (2014)	Plano de preservação
	Preservação dos dados
	Financiamento para preservação
	Custos da preservação
	Qualidade dos dados
	Atuação de profissionais da informação
Sturges et al. (2015)	Serviços de suporte
	Qualidade da comunicação científica
Childs e Mcleod (2004)	Serviços de orientação aos pesquisadores
	Treinamento de profissionais
Douglass et al. (2014)	Serviço de suporte aos pesquisadores

Fonte: Elaboração própria.

Segundo Knight (2015), a oferta do serviço para a gestão dos dados de pesquisa na University of London é um dos elementos que promovem o sucesso da comunicação dos dados de pesquisa na instituição. No entanto, apesar de sua presença, infere-se que o serviço não contempla o estágio de elaboração do plano de gestão, visto que a etapa foi apresentada como problemática pelos pesquisadores, por causa da falta de conhecimento sobre seus processos e padrões necessários (Figura 49).

Figura 49: Desdobramentos da categoria “gestão dos dados de pesquisa”.



Fonte: Elaboração própria.

O estudo de Whitmire, Boock e Sutton (2015) também é contextualizado em um serviço de gestão de dados de pesquisa ofertado na Oregon State University. Após a identificação das barreiras apontadas pelos pesquisadores para a comunicação dos dados de pesquisa, os autores concluíram que o serviço ofertado pela universidade precisa ser mais bem divulgado. Isso porque parte dos problemas apontados seriam sanados pelo serviço da universidade, caso ele fosse utilizado pelos pesquisadores.

Diferentemente dos estudos anteriores, Gardner (2004) não parte de uma experiência positiva de serviço para a gestão de dados. O autor apontou a necessidade de um serviço dessa natureza como uma solução necessária às demandas de disponibilização dos resultados de pesquisa impostas pelo FOIA. Gardner ressalta que, sem essa infraestrutura básica, que pressupõe o compartilhamento de responsabilidade para o processo, a exigência de acesso aos dados inviabilizaria o desenvolvimento de

pesquisas, especialmente aquelas que contrariam interesses industriais, políticos e comerciais.

O compartilhamento de responsabilidade também foi enfatizado por outros dois artigos. Um deles é de autoria de Childs e Mcleod (2004), que recomenda o aprimoramento dos serviços de orientação aos pesquisadores, bem como a ampliação do treinamento dos profissionais da informação envolvidos no processo. O outro estudo, conduzido por Dallmeier-Tiessen et al. (2014), apresentou argumentos novos à discussão. Segundo os resultados apresentados, os serviços de gestão de dados de pesquisa é condição necessária para conferir qualidade aos dados de pesquisa comunicados. Além disso, a qualidade dos dados somada à padronização dos registros ampliaria a possibilidade de visibilidade e citação dos dados e da pesquisa. Ambos os resultados seriam possíveis a partir da atuação de profissionais da informação qualificados. Tais profissionais seriam capazes de avaliar e normalizar os registros dos dados, contribuindo assim para a gestão efetiva dos dados de pesquisa e sua preservação em longo prazo. Apesar da relevância das práticas de preservação, os autores ponderaram que os custos e, por vezes, a falta de financiamento são empecilhos para a garantia da preservação. A valoração da atuação dos profissionais da informação no processo de preservação dos registros dos dados também foi empenhada, de forma geral, por Thompson, Robertson e Greenberg (2015).

Além dos processos relacionados com o suporte aos pesquisadores e com a preservação dos sistemas de informação, a atuação dos profissionais da informação também foi evidenciada por Cox, Verbaan e Sen (2012), como a peça-chave para a articulação da comunicação dos dados de pesquisa. Segundo os autores, esses profissionais destacam-se no contexto em decorrência de seus conhecimentos técnicos acerca dos sistemas de informação e das práticas na gestão da informação. Tal expertise ainda é relevante para proposição e consolidação das políticas de informação em uma instituição.

4.2.4.4. Políticas de informação

As políticas de informação para a comunicação dos dados de pesquisa foram citadas por quatro artigos, a saber: Dallmeier-Tiessen et al. (2014), Melero e Hernández-San-Miguel (2014), Gardner (2004) e Sturges et al. (2015) (Quadro 33).

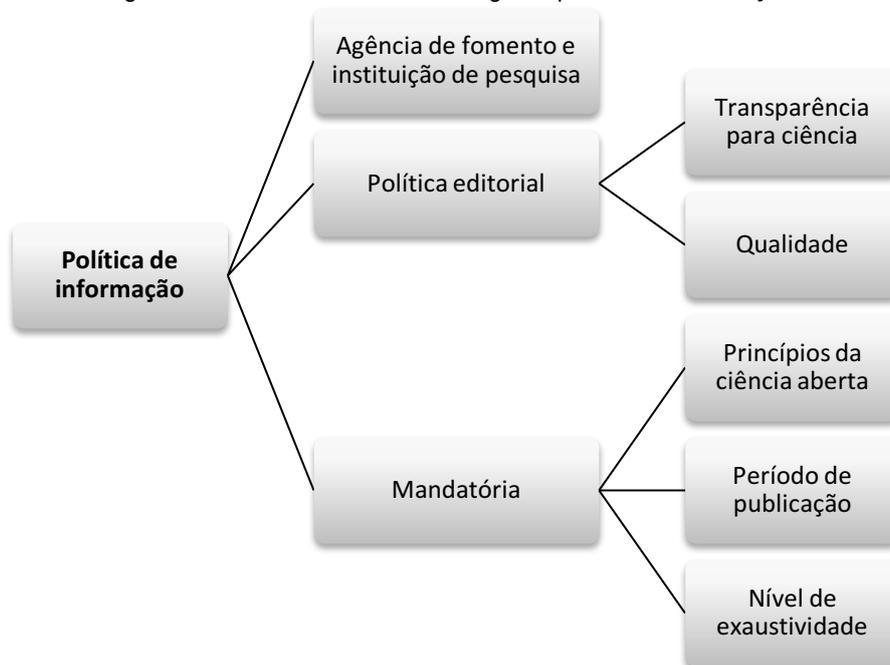
Quadro 33: Políticas de informação.

Autores	Codificação dos dados
Gardner (2004)	Tempo para tornar os dados disponíveis Divulgação exaustiva dos dados
Melero e Hernández-San-Miguel (2014)	Políticas de acesso aberto
Cox, Verbaan e Sen (2012)	Políticas de informação
Dallmeier-Tiessen et al. (2014)	Políticas mandatórias
Sturges et al. (2015)	Políticas editoriais

Fonte: Elaboração própria.

Para Dallmeier-Tiessen et al. (2014), as políticas que requerem, e não apenas sugerem, a divulgação dos dados de pesquisa são incentivos estratégicos para o sucesso da prática. Segundo os autores, as políticas precisam ser determinadas pelas instituições dos pesquisadores ou agências que fomentam suas pesquisas (Figura 50).

Figura 50: Desdobramentos da categoria “política de informação”.



Fonte: Elaboração própria.

No mesmo sentido, foi apresentado o argumento de Melero e Hernández-San-Miguel (2014). No entanto, foi enfatizado o aspecto da orientação para os princípios do acesso aberto na divulgação dos dados de pesquisa. Melero e Hernández-San-Miguel (2014) ressaltam que a forma de disponibilização dos dados de pesquisa deve seguir

as prerrogativas do acesso aberto e pressupor que ela seja feita por meio da internet e sem restrições de acesso ao material.

A perspectiva da necessidade da política de informação também parece fundamentar a discussão proposta por Gardner (2004). O autor, que problematizou a proposta do FOIA, afirmou que a política de informação para divulgação dos dados de pesquisa deve prever o período entre a finalização da pesquisa e a publicação dos dados brutos. Adicionalmente, a política deverá indicar o nível de exaustividade em que os dados poderão ser demandados.

Em todos os três casos, evidencia-se a preocupação com a política de informação como um instrumento que promove a comunicação dos dados de pesquisa, com vistas a garantir a qualidade da comunicação científica. No entanto, a relação só foi declaradamente estabelecida por Sturges et al. (2015). Os autores discutiram o papel das políticas editoriais dos periódicos científicos como requisito para garantir a transparência dos resultados das pesquisas que publicam. Nesse sentido, a transparência seria um elemento-chave para promover a qualidade dos periódicos científicos.

4.3. Conjunto das categorias preliminares, suas dimensões e propriedades

Após a identificação das categorias preliminares e de suas propriedades, foi possível classificá-las em função de uma dimensão de análise. Os três elementos aqui apresentados são pré-requisitos para a análise dos dados das entrevistas, que será discutida no próximo capítulo. Portanto, a revisão sistematizada da literatura resultou em um conjunto de 17 categorias preliminares e 61 propriedades, classificadas em 4 dimensões de análise (Quadro 34). As categorias preliminares e os seus desdobramentos representam os fatores que influenciam a comunicação de dados de pesquisa, segundo a literatura científica, cumprindo assim o primeiro objetivo específico do estudo.

Quadro 34: Categorias preliminares resultantes da RSL.

	Categoria preliminar	Vertente (propriedade)
Imperativos da ciência moderna	Aspectos éticos	Privacidade dos participantes
		Anonimização dos dados
		Permissões dos participantes e de órgãos reguladores
	Aspectos legais	Direitos de propriedade privada
		Licenças de uso
		Interesses comerciais
		Legislação
		Direito à informação
		Direito à saúde
	Função social	
	Assédio aos pesquisadores	Solicitações excessivas para divulgação dos dados
	Aceleração da ciência	Rapidez na produção de novos resultados
		Maximização no uso dos dados e na produção de novos conhecimentos
		Redução do custo na produção de novos resultados
	Cultura acadêmica	Área do conhecimento
		Interesses comerciais
		Idade dos pesquisadores
		Tecnologias de comunicação
	Recompensa ao pesquisador	Valor acadêmico
		Status e promoção na carreira de pesquisador
		Incentivos
	Reuso	Rapidez na produção de novos resultados
		Redução do custo na produção de novos resultados
Aplicabilidade em outras pesquisas		
Desqualificação dos dados e seus resultados		
Fundamentação de argumentos contrários		
Transparência da ciência	Validação dos resultados	
	Fundamentação de novas teorias	
	Integridade e compreensão científica	
Visibilidade	Identificação de erros	
	Reconhecimento do trabalho	
	Atribuição da autoria	
	Citação	
Contexto social	Saúde pública	Políticas de informação
		Ciências da saúde
		Função social
		Maximização no uso dos dados e na produção de novos conhecimentos
		Aspectos éticos
Financiamento	Custo	Infraestrutura
		Preservação em longo prazo
		Economia de recursos
		Desigualdade nos recursos para análise de dados
	Bem público	Políticas de informação
		Economia de recursos
		Democracia
	Instituição financiadora	Políticas de informação
		Economia de recursos
		Estabelecimento de requisitos

	Categoria preliminar	Vertente (propriedade)
Infraestrutura	Gestão dos dados de pesquisa	Serviços de informação
		Orientação aos pesquisadores
		Padronização dos registros
		Preservação em longo prazo
		Custo
		Profissionais da informação
		Compartilhamento da responsabilidade
		Infraestrutura
		Metadados
		Qualidade
		Visibilidade
		Citação
		Políticas de informação
		Padrão
	Interoperabilidade	
	Metadados	
	Tipos de dados	
	Estratégias de preservação	
	Política de informação	Instituição financiadora
		Política editorial
		Transparência para ciência
		Qualidade para ciência
		Estabelecimento de requisitos
	Sistema da informação	Repositório de dados
		Periódicos acadêmicos
		Sistemas de descoberta
		Segurança
		Tecnologia
		Interoperabilidade
		Divulgação

Fonte: Elaboração própria

O conjunto das categorias preliminares deram as condições para a coleta dos dados e a emergência das categorias teóricas, que serão apresentadas no próximo capítulo.

5. FATORES QUE INFLUENCIAM A COMUNICAÇÃO DOS DADOS DE PESQUISA SOBRE O VÍRUS DA ZIKA NA PERSPECTIVA DOS PESQUISADORES

A discussão sobre os fatores que exercem influência na comunicação dos dados de pesquisa é estruturada a partir da metodologia da teoria fundamentada, proposta por Strauss e Corbin (2008). Nessa proposta, a teoria emerge do desenvolvimento de três etapas: descrição da pesquisa, ordenação conceitual e o refinamento da teoria substantiva.

Na descrição, devem ser apresentados o ambiente da pesquisa, os participantes e o conteúdo resultante da entrevista. Para tanto, o conteúdo é sistematizado por meio da codificação aberta dos dados. Segundo Strauss e Corbin (2008), a codificação aberta é um “processo analítico por meio do qual os conceitos são identificados e suas propriedades e dimensões são descobertas nos dados” (p. 103). As categorias emergentes dessa etapa também foram relacionadas com as categorias preliminares, obtidas na fase anterior. No entanto, os autores ressaltam que o diálogo não deve ser executado no intuito de confirmar as categorias preliminares, e sim de expandi-las e, se necessário, refutá-las. Em decorrência da codificação aberta, os dados são reduzidos para um bloco menor e mais maleável, que é a base para a ordenação conceitual.

A ordenação conceitual representa a organização e associação de categorias, ao nível de propriedades e dimensões, por meio da codificação axial. Segundo Strauss e Corbin (2008, p. 124), “o objetivo da codificação axial é começar o processo de agrupamento dos dados que foram divididos na codificação aberta”, de forma que seja possível gerar explicações mais completas sobre o fenômeno. Portanto, todas as categorias obtidas nas entrevistas foram reunidas e reorganizadas.

Posteriormente foi realizada a codificação seletiva dos dados, que representou a integração e o refinamento da teoria. Para tanto, destacou-se a categoria central e sua relação com as demais categorias. O resultado dessa etapa corresponde à proposição da teoria substantiva, última fase do estudo.

5.1. Descrição e codificação aberta

Todos os pesquisadores foram selecionados a partir da busca na base DGP/CNPq, segundo os critérios apresentados na metodologia. Os entrevistados compuseram assim a amostra homogênea teórica. Sua homogeneidade foi alcançada

pelo nível de especialização no tema apresentado pelos pesquisadores entrevistados. Seu aspecto teórico foi obtido por meio da identificação de categorias teóricas pertinentes para a discussão. Apesar de terem sido identificados 13 pesquisadores como potenciais para entrevista, o ponto de saturação foi alcançado com a realização de apenas 10. A saturação da amostragem teórica representou a repetição das categorias já identificadas.

5.1.1. Considerações acerca da entrevista com o P01

O primeiro pesquisador entrevistado tem formação na área de ciências biológicas e atua em microbiologia e genética, com ênfase em virologia. Seu projeto de pesquisa abarca temas relacionados ao *Aedes aegypti* e suas arboviroses, tais como zika e dengue. Segundo a descrição do projeto, o grupo de pesquisa foi organizado em sete subgrupos. Cada um deles conta com um pesquisador coordenador do tema. O P01 lidera o subgrupo que trata da genética do vírus da zika e da criação de um banco de dados com informação genética.

A entrevista com o P01 ocorreu em seu local de trabalho e durou cerca de 40 minutos (Apêndice C). O pesquisador pareceu familiarizado com o tema e favorável às práticas de comunicação aberta dos dados de pesquisa. Sua principal motivação é a promoção do reuso dos dados, de forma a contribuir para o avanço da ciência. Segundo o pesquisador, tais práticas fazem com que o resultado da pesquisa seja mais completo e produza publicações com melhor impacto.

Principalmente a reutilização, porque eu venho de uma área de pesquisa que eu dependo muito desses bancos de dados que as pessoas criam. Porque eu trabalho com evolução de vírus, então, eu preciso do dia que a amostra foi coletada, o local, o hospedeiro. Eu sofro muito com esse tipo de informação, porque coisas antigas não têm esse tipo de coisa ou, às vezes, os autores, eles não colocam ou por preguiça ou porque eles não têm mais esses dados, de que dia que foi coletado, esse tipo de coisa. Então, eu já venho me envolvendo com esse tipo de coisa, porque eu preciso disso. [...] Mas eu acho que a vantagem é essa: se eu fizer isso e outra pessoa fizer, outra fizer, todo mundo ganha no final, porque a gente vai conseguir esses dados juntos e no final das contas acho que isso gera muita colaboração. Se eu tenho essa informação, outra pessoa tem outra informação, outra pessoa tem outra informação, às vezes, o que quase sempre acontece, as pessoas juntam num único artigo grande.

O pesquisador considera que essa prática de colaboração científica é recorrente em seu campo de pesquisa. Para sustentar seu ponto, ele exemplifica:

Mas se esse compartilhamento ocorre, as pessoas vão trocando informação, no final junta tudo isso num artigo um pouco mais completo. É o que eu acho que vai acontecer agora: vai ter um artigo grande de origem e evolução da zika. Vai envolver pesquisadores da América Latina inteira, pesquisadores dos Estados Unidos, pesquisadores da Europa. Está todo mundo fazendo

vinte genomas, outros tão fazendo vinte, aí eles vão juntar isso tudo em um único artigo. Em um único artigo grande de impacto.

O reuso dos dados provocaria outros efeitos positivos, tais como a economia de recursos e a identificação de possíveis erros da pesquisa. Apesar das vantagens e do posicionamento favorável, foram pontuados cinco aspectos que dificultam a comunicação dos dados de pesquisa: as diferenças regionais, especialmente a falta de cultura de colaboração científica no Brasil; a desvantagem internacional na colaboração; a necessidade de autorização para o uso dos dados; o idioma e os esforços para o tratamento dos dados.

Os dois últimos aspectos foram especialmente destacados como empecilhos para a prática. Segundo o pesquisador, o tratamento dos dados que serão divulgados demanda esforços adicionais para seu registro e organização. Além disso, os dados precisam ser registrados e descritos necessariamente em inglês.

[...] eu já teria que organizar isso tudo em um idioma: o inglês. Então, todos os meus protocolos, todos os meus dados já teriam que estar organizados em inglês. Se eu não fiz desde o início, então eu vou ter que ter esse... se eu quiser compartilhar, eu vou ter que ter essa etapa pra poder colocar tudo em inglês. Então, isso seria um trabalho a mais, que eu teria que fazer, pra poder compartilhar. Ou você já decide que você vai organizar toda a informação em inglês desde o início. Esse é um empecilho.

O outro empecilho destacado pelo pesquisador trata da ausência de infraestrutura adequada dentro da universidade para disponibilizar os dados.

[...] eu não preciso disponibilizar meu banco de dados, eu poderia disponibilizar em alguma coisa aqui dentro da universidade mesmo, que a universidade, ela tivesse uma infraestrutura (às vezes até tenha, a gente desconheça) pra você poder fazer esse tipo de coisa de uma maneira mais simples, sabe, aonde você pudesse alimentar isso, não sei, acho...

Além de benefícios e empecilhos, o pesquisador citou aspectos que influenciam a prática, podendo tanto promovê-la como inibi-la. Tais aspectos foram categorizados, por meio da codificação aberta, em: idade dos pesquisadores, possibilidades de acesso aos dados, políticas editoriais das publicações, licenças de uso e contexto de emergência da saúde pública (Quadro 35).

Quadro 35: Codificação aberta da entrevista - P01.

Dimensão	Categoria preliminar	Categoria	Propriedade
Imperativos da ciência	Acionamento da ciência	Acionamento da ciência	Produção de novos resultados
	Aspectos legais	Autorização para uso dos dados	Solicitada diretamente para os pesquisadores
	Reuso	Colaboração	Necessário para o reuso dos dados
	Cultura acadêmica	Idade dos pesquisadores	Determina comportamento
	Transparência para ciência	Identificação de erros	
	Reuso	Reuso	Produção de novos resultados
Contexto social da doença	Saúde pública	Contexto da saúde pública e das doenças emergentes	Emergência de saúde pública
Financiamento	Custo	Desvantagem internacional na colaboração	Desigualdade de acesso aos recursos e diferenças regionais
	Custo	Economia de recursos (tempo e esforços)	Custo
Infraestrutura	Sistema de informação	Banco de dados	Dados sobre genoma
	Sistema de informação	Políticas editoriais das publicações	Requisitam os dados
Fenômeno		Idioma no tratamento dos dados	Inglês
		Banco de dados internacionais	Padrões internacionais
		Acesso aos dados brutos	Prática comum

Fonte: Elaboração própria.

5.1.2. Considerações acerca da entrevista com P02

A segunda pesquisadora entrevistada também tem formação na área das ciências biológicas e atua em microbiologia e genética molecular humana, com ênfase em virologia. Entretanto, sua linha de pesquisa difere da abordagem de P01, seu estudo abrange investigação genética sobre formação cerebral associado ao vírus da zika.

A entrevista com P02 ocorreu em seu local de trabalho e durou cerca de 30 minutos (Apêndice D). A pesquisadora esclareceu que trabalha com duas naturezas de dados e, para cada uma delas, foi apresentada uma perspectiva diferente. Um dos tipos origina-se da genética molecular. Para esses dados, P02 afirmou que utiliza dados primários compartilhados por meio de um banco de dados. Para a pesquisadora, o sistema promove maior celeridade para as descobertas científicas.

“Não, se não fosse esse banco, a gente ia tá dez mil anos atrás. Não tem como, tem que ter.”

Para os dados da genética molecular, não foram apontados empecilhos para sua comunicação aberta. No entanto, para o outro tipo de dados, que diz respeito aos dados dos pacientes, foram apresentados seis fatores que dificultariam sua comunicação aberta, são eles:

4. perda de ineditismo;
5. diferenças de recursos e capacidades de análise dos dados entre os pesquisadores;
6. aspectos éticos;
7. aplicabilidade dos dados anonimizados;
8. ausência de orientação dos periódicos científicos;
9. ausência de banco de dados.

Entre as barreiras apontadas por P02, considerou-se que a principal é a perda de ineditismo. A possibilidade de outro pesquisador utilizar seus dados para desenvolver uma pesquisa e comunicá-la primeiro que o pesquisador que coletou os dados está diretamente relacionada com outras barreiras: as diferenças de acesso aos recursos necessários à pesquisa e capacidades de análise dos dados.

E o que eu vejo muito, nesses anos todos, é que, às vezes, você tem uma ideia e outro do outro lado ou aqui mesmo do meu lado, na sala ao lado, teve a mesma ideia, só que ninguém compartilha isso. Então você vai... acaba que você vai fazer a mesma coisa e isso invalida na hora de você publicar, porque não é mais inédito, aí o outro, porque tem uma equipe maior, que tem mais recurso financeiro, publica na sua frente. Então, eu acho que essa informação, do que que você tá fazendo. [...] Então, eu acho que essa informação, do que que você tá fazendo, eu acho que é interessante, mas perigosa. Perigosa nesse sentido, de perder ineditismo.

Apesar das barreiras, P02 considerou três benefícios para a ciência:

10. evitar o retrabalho e economizar recursos;
11. acelerar a produção de novos resultados;
12. melhorar a qualidade de vida da população;

O último benefício citado relaciona-se diretamente com a área do conhecimento das ciências da saúde. Pela sua importância, a pesquisadora considerou que esse benefício a motivaria a comunicar abertamente seus dados.

O que me motivaria seria isso, acho que o motivo, né, quer dizer, é você contribuir com a ciência para que se chegue a uma resposta mais rápida para o paciente. Para melhorar a qualidade de vida do paciente. Acho que esse é o objetivo da ciência: melhorar a qualidade de vida do indivíduo, do paciente no caso.

Em relação às suas práticas de produção, compartilhamento e uso de dados relacionados aos pacientes, P02 declarou não utilizar nenhum sistema de comunicação aberta, apesar de o grupo possuir um banco de dados resultantes de suas pesquisas.

P02: ... é o que a gente faz aqui. O laboratório, por exemplo, ele tem um banco de dados... Mil e trezentos e alguma coisa de pacientes [identificação do banco], que é o carro-chefe do laboratório. A partir desse banco que nós montamos, a gente fez, sei lá, cinco, seis, dez projetos em cima desse mesmo dado. Só que investigando uma outra coisa, né? Mas utilizando os mesmos dados do paciente. Entende o que eu tô falando?

MC: Esses dados, eles ficam aqui disponíveis para os pesquisadores aqui da [Instituição A] ou...

P02: Não. Dentro do grupo de [pesquisa X], sim.

MC: Daqui da [Instituição A]?

P02: Não da [Instituição A]. De dentro no [Instituição B].

A falta de abertura dos dados parece ser resultante da ausência de políticas e sistemas para tal finalidade. A dedução aqui apresentada baseia-se em outra declaração de P02 em relação às experiências internacionais que condicionam o financiamento da pesquisa à publicização dos dados. A pesquisadora P02 considerou que, para a aplicação de políticas similares no Brasil, seriam necessárias ofertas de infraestrutura adequada e serviços de informação.

Eu acho que para gerar isso tem que se criar, é aquilo que eu falei, uma central de dados, onde tivesse gerentes dessa central de dados, pra fazer tudo direitinho. Porque, assim, no Brasil, por exemplo, no nosso laboratório: eu faço tudo, a gente faz tudo então, eu sou a pesquisadora, chefe do laboratório; eu tenho que escrever o projeto; eu tenho que veicular os dados; eu tenho que ver se os dados estão certos; eu tenho que ir pra bancada ver se tá bom; eu tenho que fazer tudo. Então, é impossível você criar um banco direitinho e disponibilizar isso.

É possível perceber no argumento de P02 que, além da estrutura, é esperado que a responsabilidade com a comunicação dos dados seja compartilhada com outros atores da comunicação científica. Tal argumento foi defendido por Gardner (2004), que ponderou que a obrigatoriedade da divulgação ampla dos dados sem políticas e infraestrutura adequada pode inviabilizar o desenvolvimento da pesquisa.

Os principais fatores apresentados por P02 foram categorizados por meio da codificação aberta (Quadro 36).

Quadro 36: Codificação aberta da entrevista - P02.

Dimensão	Categoria preliminar	Categoria	Propriedade
Imperativos da ciência	Cultura acadêmica	Tipos de dados	Dados de genética molecular
	Reuso	Aplicabilidade dos dados	Dados dos pacientes
	Reuso	Ineditismo	Dados anônimos
	Aspectos éticos	Retrabalho	Ego dos pesquisadores
	Aspectos éticos	Aspectos éticos	Economia de recursos
Contexto social da doença	Aceleramento da ciência	Aceleramento da produção de novos resultados	Privacidade dos pacientes
	Saúde pública	Melhoramento da qualidade de vida das pessoas	Emergência de saúde pública
Financiamento	Custo	Capacidade de análise	Objetivo da ciência
	Instituições	Compartilhamento da responsabilidade	Falta de recursos necessários
Infraestrutura	Sistema da informação	Políticas editoriais	Responsabilidades
	Sistema da informação	Banco de dados	Publicações acadêmicas
			Acesso controlado

Fonte: Elaboração própria.

5.1.3. Considerações acerca da entrevista com P03

A terceira pesquisadora entrevistada tem formação na área de farmácia, ciências biológicas e biotecnologia. Sua área de atuação é a virologia, com ênfase em epidemiologia. Sua linha de pesquisa abrange estudo sobre a atividade antiviral do vírus da zika para o desenvolvimento de vacina.

A entrevista com P03 foi realizada em seu ambiente de trabalho e durou cerca de uma hora. Em decorrência de sua recomendação, também foi entrevistado outro membro do seu grupo de pesquisa, que trabalha diretamente com a coleta e o registro dos dados. A entrevista com a outra pesquisadora foi realizada na mesma ocasião e será descrita no próximo item desta seção.

De forma geral, P03 se posicionou de forma contrária à comunicação aberta dos seus dados de pesquisa. O principal problema apontado diz respeito ao sigilo da pesquisa, especialmente em razão da natureza da sua área.

Olha só, o que eu posso dizer é que normalmente na minha área, que é uma área muito competitiva, ninguém abre a boca pra falar nada. As pessoas não abrem muita coisa do que é o segredo de estado, tá? [...] Então, em alguns casos ou alguns professores, eles que fecham esses dados, porque eu trabalho mais com iniciação tecnológica, produto que possa gerar o fármaco que possa lá na frente vir a ser o medicamento, ou fármaco ou medicamento, que já esteja numa farmácia. Então, isso aí existe uma proteção, então a gente não fala muito, tá?

A pesquisadora P03 apresentou uma nova perspectiva para a discussão, a partir do ponto de vista de quem está diretamente produzindo um fármaco. Nesse contexto, a abertura dos dados não é interessante para sua prática de pesquisa. Outros fatores foram destacados em complemento à necessidade de sigilo da pesquisa, especialmente a desigualdade na capacidade tecnológica para processar os dados, o que permitiria chegar mais rapidamente aos resultados desejados.

[...] a gente não pode falar... essa ideia que a gente tem, que está na cabeça, não pode sair, porque se não a gente que é pobre, a gente vai dar ideia pros outros, que tem muito dinheiro [...] Faz mais rápido e a gente quando vê a sua ideia já foi feita. [...] Geralmente a gente não acha interessante dizer o que tá na cabeça pra outras pessoas, porque ela pega tua ideia, como ela tem dinheiro, tem gente pra fazer, maior número de gente pra fazer, e ela faz isso com mais rapidez do que você que não tem esse dinheiro.

A falta de recursos para a análise dos dados foi igualmente destacada por P03 no processo de produção desses dados. Os principais dados produzidos pelo grupo de pesquisa foram financiados com recursos dos próprios pesquisadores.

É difícil financiamento. Muita coisa saiu do meu bolso. [...] Mil novecentos com os prime, sonda... eu paguei do meu bolso.

Além de terem de financiar parte da pesquisa com recursos particulares, os dados que comunicam para outros projetos de pesquisa não resultam em reconhecimento e/ou recompensas para os pesquisadores que coletaram os dados.

Eu passo os dados e nem sempre os dados que eu passo eu sou premiada com o nome no paper. Ele [...] bota lá o valor que eu dei, como se fosse eles que fizeram. [...] Entendeu? Eu já vi isso, eu já passei resultado citotoxicidade, depois as pessoas publicaram, botaram o resultado que eu fiz, eu gastei tempo, gastei material, aí agora quando eu faço esse tipo de trabalho pra alguém, porque às vezes eles são enrolados com o tempo em termo de célula, eles não conseguem manipular direito a célula, aí mata a célula, contamina a célula, aí perde a célula, aí eu digo: "eu faço, mas com uma condição: o nome tem que tá no paper".

Dadas as condições específicas do contexto de P03, sua prática de comunicação dos dados de pesquisa é essencialmente fechada. As preocupações com o controle e o sigilo da pesquisa são facilmente identificadas em todos os macroprocessos da comunicação dos dados. Segundo P03, o registro dos dados se dá de forma híbrida, em formato digital e analógico. Em ambos os casos, existe a preocupação em registrar os dados de forma codificada para dificultar o entendimento de outras pessoas que não são convidadas a colaborar com a pesquisa.

Tudo meu tá no computador, existe um computador meu, o meu aluno tem o dele, o outro tem o dele e tem esse aqui, todos os computadores têm os resultados. Com o seguinte, porque se esse perder o computador de um tem o resultado do computador do outro, porque a informática permitiu isso. Mas eu sou mais ainda... assim, eu penso que o papel ainda existe e, aí, esse não vai apagar. Então, eu tenho no meu setor, eu trabalho com zika brasileiro, caderno de zika brasileiro com os resultados [...] O caderno de zika africano com os meus resultados, o que me interessa ou então tá em código, porque se tiver em código ninguém vai saber o que que é.

O compartilhamento dos dados também é feita de modo que alguns dados possam ser preservados para uso exclusivo do grupo de pesquisa. Em geral, a forma de disponibilização dos dados ocorre por meio da publicação de artigos científicos, nos quais os resultados e os dados relevantes para a discussão do tema são publicados. No entanto, os resultados e seus dados considerados como “muito interessantes” são patenteados.

[...] algumas coisas a gente publica e o que a gente achar que é muito, muito interessante a gente patenteia, porque o projeto da gente é biotecnologia [...]

O contexto não propício para a produção e compartilhamento aberta dos dados resulta, ou é consequência, da ausência de práticas de reuso de dados de outras pesquisas. Segundo P03, os dados resultantes de outras pesquisas não são aplicáveis ou interessantes para o seu projeto de pesquisa em decorrência das especificidades demandadas.

MC: E pra produzir, a senhora e seu grupo, tem prática de utilizar dados de outras pesquisas?
 P03: Eu não, eu não confio. Se tiver que fazer, eu repito tudo de novo. [...] Eu repito tudo de novo. Se eu pegar uma droga lá em cima, disse: “ah, já tem cento e cinquenta”. Eu to trabalhando agora com o zika, replica em vero, eu não trabalhava tanto com a vero. [...] não é que tá errado lá, não é que eu vou estar errada, nem que a pessoa que fez lá tá errada, não é isso. São momentos diferentes de uma célula. Pode ser uma célula mais nova, bombando e, de repente, você tá com uma célula mais velha, que ela ainda está funcionando, mas a gente não sabe como que ela tá se comportando depois de tanta passagem, de tanto estresse.

A perspectiva apresentada por P03 apresentou novos elementos para a análise geral do fenômeno. Os principais fatores apresentados por P03 foram categorizados por meio da codificação aberta (Quadro 37).

Quadro 37: Codificação aberta da entrevista - P03.

Dimensão	Categoria preliminar	Categoria	Propriedade
Imperativos da ciência	Cultura acadêmica	Sigilo da pesquisa	Inovação tecnológica Fármaco Registro dos dados

	Recompensa	Falta de recompensas e reconhecimento	Reuso Citação Publicações científicas
Financiamento	Custo	Desigualdade na capacidade tecnológica para processar os dados	Recursos
	Bem público	Ausência de financiamento público	Financiamento particular
Infraestrutura	Sistema de informação	Aspectos legais	Patente

Fonte: Elaboração própria.

5.1.4. Considerações acerca da entrevista com P04

A quarta pesquisadora entrevistada tem formação na área das ciências da saúde e experiência em epidemiologia genética e biologia computacional. Ela é integrante do grupo de pesquisa liderado por P03, que busca desenvolver um fármaco contra o vírus da zika. A entrevista com P03 foi realizada em seu ambiente de trabalho e durou cerca de meia hora.

Apesar de P04 ser do mesmo grupo de pesquisa de P03, suas percepções sobre a abertura na comunicação dos dados de pesquisa são diferentes. Conforme relatado anteriormente, P03 apresentou um conjunto de fatores que a impedem de comunicar livremente seus dados. Nem mesmo o contexto da emergência da saúde pública relacionada ao vírus da zika foi apontado com aspecto que promoveria tal abertura. Já P04 apontou alguns elementos que a motivariam a comunicar abertamente seus dados, embora também tenha sinalizado aspectos impeditivos. Os elementos que promovem a prática são:

- expectativa de avanço da ciência em menor prazo;
- validação dos resultados publicados;
- economia de recursos;
- colaboração internacional.

Sobre alguns desses aspectos a pesquisadora declarou:

[...] se a gente tivesse de alguma forma, informações, um background genético do vírus ou da população, tudo mais, isso, a gente avançaria em pesquisa, assim, anos luz, assim. Eu digo isso porque, por exemplo, lá fora a gente tem os grandes bancos de dados genéticos. Lindo. É lindo. Lá fora você consegue ver isso. Dados populacionais, dados do próprio vírus [...] Você tem essa disponibilidade pro público dessa informação, mas você também tem uns centros em que essa informação fica restrita àquele centro ou àquele país ou àqueles grupos de pesquisadores. Você imagina se a gente tivesse, se os grupos de pesquisa que tão trabalhando com dados genéticos do vírus, no genoma, na metagenômica. Você imagina se esses grupos resolvem abrir os seus resultados, criasse uma rede, uma... a gente ia, agora, por exemplo [...] Seria maravilhoso... Por

exemplo, se tem ali... vamos ver, duas pessoas testaram essas drogas na vero duas vezes, então, ali você tem um referencial de como que está aquela ação em dois períodos diferentes.

Em sua declaração, observa-se que a existência de um banco de dados é um fator condicionante para o cenário almejado. A pesquisadora ressalta que tal sistema deve ser equipado de mecanismos que garantam o controle de acesso aos recursos lá disponibilizados. A preocupação é justificada pelas diferenças de acesso aos recursos para análise dos dados. Conforme foi discutido na RSL, a partir do texto de Melero e Hernández-San-Miguel (2014) e dos argumentos apresentados por Sankoh e Ljsselmuiden (2011), a falta de acesso aos recursos necessários para analisar os dados com rapidez e eficiência colocaria os países e regiões em desenvolvimento sempre atrasados em relação a comunicação de novos resultados.

[...] Com outro grupo lá fora, porque tem isso já estabelecido no laboratório dele. Então, é mais fácil do que eu montar aqui. É mais barato eu ficar lá seis meses e trazer essas modificações pra depois voltar e trabalhar aqui. Porque montar aqui numa época que tá pobre o país, é complicado. Eu não teria dinheiro pra montar essa estrutura que eles têm lá já montado. [...] Que exige uma capacidade de computador... Que aqui não tem. Não, nem consegue. O dela ali tá capenga, faz em três horas o que o outro faz em um minuto.

A pesquisadora ressalta que faltam às regiões mais pobres recursos de infraestrutura necessários para analisar os dados. No entanto, tais regiões, por vezes têm acesso privilegiado aos dados necessários para o desenvolvimento de uma pesquisa.

Só que, às vezes, a ideia que a gente tem, eles não pensaram. Sim, e muitas vezes até os acessos aos dados, ao material biológico que a gente tem aqui... A gente pensa, tá a gente pensa. Tá pensando o que, a gente pensa e muito. Só não tem dinheiro. Infraestrutura. A gente até que... essa parte, que como é pra fazer, rodar genoma, a gente não tem mesmo, tá? A gente não tem.

Somada à ausência de infraestrutura e às desigualdades no acesso aos recursos para análise dos dados, a falta de recompensas acadêmicas pela divulgação dos dados brutos também foi apresentada como um fator inibidor da prática. A pesquisadora P04 afirmou que os pesquisadores, de modo geral, são cobrados pela produtividade, medida especialmente pela publicação de artigos e criação de patentes. Com isso, a ampla divulgação dos dados coletados para uma pesquisa não gera valor acadêmico para os pesquisadores e ainda pode prejudicá-los. O efeito maléfico seria decorrente da perda de sigilo da pesquisa e da possibilidade de produzir novas publicações inéditas com os dados coletados.

As práticas de produção, compartilhamento e uso dos dados por P04 são as mesmas de P03. No entanto, foi identificada uma diferença em relação à perspectiva de ambas. P04 mostrou-se mais disposta para a abertura dos dados do que P03. Dois fatores apontados pela RSL e por P01 podem ajudar a explicar essa diferença. Um deles é a idade dos pesquisadores, uma vez que P04 é mais jovem do que P03. O outro elemento é a área do conhecimento. A formação de P03 relaciona-se com a área de Farmácia, enquanto a de P04 está voltada para Biologia.

As categorias e propriedades resultantes da análise da entrevista de P04 foram descritas no Quadro 38.

Quadro 38: Codificação aberta da entrevista -P04.

Dimensão	Categoria preliminar	Categoria	Propriedade
Fenômeno		Acesso a outros dados	Falta de dados das pesquisas sobre o vírus da zika
Financiamento	Custo	Desigualdades no acesso aos recursos para análise dos dados	Relação dos países em desenvolvimento (produtores de dados) e países desenvolvidos (maior capacidade de análise)
	Custo	Economia de recursos	Evita trabalhos duplicados
Imperativos da ciência	Cultura acadêmica	Sigilo da pesquisa	Recurso das regiões com menos capacidade de análise
	Recompensa	Mecanismos de avaliação de produtividade a partir dos dados comunicados	Aspecto da cultura que inibe a comunicação
	Cooperação internacional		Provocado pela necessidade de compartilhar recursos
	Aceleração da ciência	Aceleração da ciência	Premissa baseada nas experiências dos bancos de dados genéticos
Infraestrutura	Transparência da ciência	Validação dos resultados	Citação de outros trabalhos
	Sistema de informação	Banco de dados	Colaboração nacional com restrição de acesso

5.1.5. Considerações acerca da entrevista com P05 e P06

O quinto pesquisador entrevistado tem formação na área de medicina e microbiologia, área na qual atua há cerca de 40 anos. Por recomendação de P05, outro pesquisador da sua equipe participou da entrevista para complementar as informações acerca do processo de coleta e tratamento dos dados. P06 também tem formação em microbiologia e no projeto de pesquisa relacionado ao vírus da zika, no qual está atuando no processo de tratamento dos dados. Já P05 é o líder do grupo de pesquisa e atua no gerenciamento de diversas frentes do laboratório.

Ambos os pesquisadores demonstraram ter conhecimento sobre as discussões acerca da ciência aberta e se posicionaram favoráveis a elas. Entretanto, P05 declarou ter tido os primeiros contatos com a ideia nos anos 1980, período em que trabalhou em um órgão de saúde pública nos Estados Unidos. No período, a discussão foi levantada em decorrência da proposta de implementação de patente para os resultados de pesquisa produzidos pela instituição. Segundo P05, houve uma resistência à proposta, uma vez que há o entendimento de que a patente é um fator que dificulta a circulação dos dados e a produção de novas informações.

[...] eu sou da época antes de patente. Eu trabalhei muito tempo em órgão de saúde pública nos Estados Unidos, no Centers for Disease Control, que é um órgão de saúde pública americano. Quando eu trabalhei lá, até 1986, a gente discutia muito isso naquela época. Quando começaram a surgir as ideias de patentes, quer dizer, as informações não serem expostas. Nós debatemos muito isso e o governo americano na época tinha uma posição muito legal, eu achava. A grande parte das pesquisas que eram feitas lá eram para ajudar a população mais pobre. Então, a patente criava um problema, a não informação de dados novos, dados que você estava desenvolvendo. Por que o que acontece, você desenvolvendo um dado, eu posso dar pra uma outra pessoa pegar e progredir com aquele dado e mais a frente ter mais condições de trabalho.

Entre as problemáticas apresentadas por P05, destaca-se a finalidade das pesquisas, que era a melhoria da qualidade de vida das populações pobres. Com isso, pressupõe-se que não cabia expectativa de lucro com os resultados da pesquisa. Seus resultados deveriam sim atender a uma função social. Para P05, essa noção ainda orienta sua atuação como pesquisador.

[...] toda vez que eu faço palestras perguntam: "Ah, sua patente, você vai patentear isso?" Não, não vou patentear. Se alguém quiser, pode copiar e fazer. Porque às vezes você tem um dado e a patente não é tudo. Você pode fechar e patentear aquilo e depois a empresa não vai ter interesse em desenvolver aquilo. Vai acabar desenvolvendo outra coisa.

No argumento dos pesquisadores, está implícita a perspectiva da livre circulação dos dados e dos resultados da pesquisa como elemento promotor do aceleração da ciência. Entretanto, apesar do posicionamento favorável dos pesquisadores, suas práticas de comunicação de dados não utilizam sistemas ou serviços abarcados pelas propostas da ciência aberta. Os principais canais de divulgação dos dados que produzem são os próprios artigos de periódicos e as comunicações em eventos científicos. Tais sistemas de comunicação nem sempre são adequados para a divulgação do conjunto total dos dados brutos, que, por vezes, são tolhidos pelos processos de avaliação dos artigos.

O que acontece é que você manda para uma revista e a revista pode achar que aqueles dados não são tão importantes. No meu ponto de vista é, mas não quer dizer que... O que que acontece, manda para seus pares, pro revisor. Dependendo do grau de entendimento do revisor.

Os dados das pesquisas de P05 e P06 são gerados a partir de amostras biológicas de pacientes infectados ou com suspeitas de contaminação pelo vírus da zika. Os dados gerados são registrados em um caderno de nota e também são armazenados nos computadores dos próprios pesquisadores.

Embora a relevância da disponibilização aberta tenha sido enfatizada pelos pesquisadores, ambos declaram não utilizar dados provenientes de outras pesquisas. A exceção restringe-se apenas às metanálises. O grupo de pesquisa de P05 e P06, realizou uma análise sistemática das publicações com resultados de pesquisa sobre o vírus da zika. Tal estudo produziu novos conhecimentos sobre o tema a partir da reunião e análise dos dados de outras pesquisas. Um dos resultados da análise é que as publicações sobre o tema são majoritariamente oriundas dos Estados Unidos.

O interesse e o domínio do país sobre o tema foram constatados pelo pesquisador P06, bem como sua necessidade de acesso aos dados coletados em regiões com grande incidência da doença, tal como o Brasil. No entanto, diferentemente das perspectivas dos outros pesquisadores, P05 e P06 não consideram que tais aspectos justifiquem a restrição do acesso aos dados ou a expectativa de exploração exclusiva dos dados pelos pesquisadores brasileiros.

[...] Inclusive o Brasil está se enfiando num buraco terrível com o negócio da zika. Porque tem pesquisador que acha que a zika é nossa, acha que temos que fazer a pesquisa. Então, começou a criar uma competição. Os americanos, os europeus, os ingleses estão a quinhentos mil anos a nossa frente. Eles terem informações dos pacientes nossos ajudaria muito a eles desenvolverem, mas felizmente tá dando essa troca. A Dilma que fez aquela proibição na época, né? De você não poder fornecer material nenhum pros Estados Unidos, fora do Brasil. Aí os pesquisadores fizeram essas colaborações um pouco meio a rebelia, mas então a não informação atrasa a pesquisa.

Na fala de P05, citada anteriormente, são apontados dois fatores. Um deles diz respeito à falta de interesse de pesquisadores brasileiros em disponibilizarem seus dados para pesquisadores estrangeiros, em especial os de países ricos com melhores recursos para análise dos dados. O outro aspecto relaciona-se com a recente Lei da Biodiversidade (nº 13.123), que entrou em vigor no Brasil em novembro de 2015.

A nova legislação revogou uma medida provisória de 2001, que regulava o acesso ao patrimônio genético para fins de pesquisa no país (BRASIL, 2015a). A partir

de então foi se estruturando um impasse jurídico, conforme explica Pierro (2016): “Sem a regulamentação e com o antigo arcabouço legal revogado, criou-se uma situação incomum, em que instituições de pesquisa e empresas ficaram sem amparo legal para obter autorizações”. Em síntese, a lei proíbe o envio para o exterior de amostras genéticas coletadas no Brasil, inclusive para finalidades de pesquisa acadêmica.

As mudanças na legislação já haviam afetado outros segmentos da pesquisa brasileira. Segundo Pierro (2016), pesquisas que dependiam de laboratórios estrangeiros, especialmente dos Estados Unidos, para o sequenciamento genético de material biológico foram paralisadas. Além disso, pesquisadores brasileiros no exterior, inclusive aqueles participantes do programa do governo federal Ciência sem Fronteiras, tiveram seus projetos de pesquisa prejudicados.

Em janeiro [de 2016], Luís Fábio Silveira, do Museu de Zoologia da Universidade de São Paulo (USP), recebeu um e-mail do Ibama que negava uma solicitação do pesquisador para enviar amostras de tecidos a um laboratório nos Estados Unidos, onde seria feito o sequenciamento genético do material biológico de aves. Como justificativa, o Ibama explicou estar “sem amparo legal no que diz respeito aos procedimentos de emissão de autorização de exploração para material biológico com finalidade de pesquisa. Diante do exposto, até que seja publicado o decreto regulamentador, estamos impossibilitados de emitir qualquer autorização de exportação que tenha como finalidade a pesquisa científica”, informou o e-mail. Silveira afirma que o entrave pode atrasar pesquisas que precisam dos resultados dos sequenciamentos para serem concluídas, prejudicando também alunos bolsistas envolvidos nesses estudos. “Interromperam, do dia para a noite, toda a nossa pesquisa. É inadmissível que não se tenha pensado em regras para a transição de uma legislação para outra”, critica. O pesquisador diz ainda que alunos que estão no exterior com bolsa do programa Ciência sem Fronteiras também dependem de amostras para a conclusão de suas teses. “Eles não têm ideia de quando isso vai ocorrer” Pierro (2016).

A problemática aqui exposta afetou diretamente as práticas de envio de dados das pesquisas brasileiras sobre o vírus da zika para o exterior. Segundo a notícia vinculada pelo portal de notícias G1, em fevereiro de 2016, cientistas estrangeiros denunciaram a falta de acesso aos dados brasileiros para o desenvolvimento de pesquisas que visam à superação das doenças causadas pelo vírus (BRASIL SONEGA, 2016). Na notícia intitulada “Brasil sonega amostras de zika para pesquisa no exterior, dizem cientistas”, fica evidente que o conflito se deu especialmente com os cientistas dos Estados Unidos e da Europa.

OMS pede que país envie mais material biológico para institutos de ponta. Centros de diagnóstico fazem exportação extraoficial de ampolas à Europa. O Brasil não está compartilhando amostras e dados o suficiente sobre o zika para permitir que pesquisadores determinem se

o vírus está realmente por trás dos casos de microcefalia, afirmam autoridades dos EUA e da Organização das Nações Unidas (ONU) (BRASIL SONEGA, 2016).

Segundo a reportagem, uma autoridade americana, que não teve sua identidade revelada, afirmou que a ausência de amostras de material biológico e dados dos pesquisadores brasileiros dificultaram a determinação dos fatores implicados na contaminação do vírus da zika e sua relação com a microcefalia. Portanto, diante de tal necessidade, a OMS teria declarado o vírus uma emergência internacional de saúde pública.

A perspectiva apresentada por P05 e P06 apresentaram novos elementos para a análise geral do fenômeno, em especial para a discussão acerca do compartilhamento de dados com pesquisadores de países desenvolvidos. Os principais fatores apresentados pelos pesquisadores foram categorizados por meio da codificação aberta (Quadro 39).

Quadro 39: Codificação aberta da entrevista - P05 e P06.

Dimensão	Categoria preliminar	Categoria	Propriedade
Contexto social da doença	Zika	Microcefalia e transmissão sexual	Preocupação internacional
	Saúde pública	Condições sociais da doença	Região tropical e serviço de saúde pública
Fenômeno		Dados dos pacientes	Recolhidos em clínicas e hospitais vinculados à instituição
		Computador próprio, caderno de notas	Dificuldade em registrar todos os dados necessários devido ao volume de trabalho
		Artigos e eventos científicos	Publicações limitam os dados divulgados
		Reuso	Não reusa dados de outras pesquisas, apenas para metanálise
Financiamento	Custo	Distribuição dos recursos da ciência	Noção de globalização
Imperativos da ciência	Reuso	Colaboração	Concorrência com os EUA
	Aspectos éticos	Questões éticas	Atraso das pesquisas
	Aspectos legais	Legislação brasileira	Proibição do envio de amostras genéticas brasileiras pra o exterior
Infraestrutura	Sistema de informação	Patente	Empecilho ao desenvolvimento da ciência de maneira mais rápida
	Sistema de informação	Artigos de periódicos	Demora para comunicar os dados e novas informação

Fonte: Elaboração própria.

5.1.6. Considerações acerca da entrevista com P07

O sétimo pesquisador entrevistado é da área de medicina veterinária com foco em microbiologia, virologia e sequenciamento de genoma. Seu projeto de pesquisa busca o desenvolvimento de alternativas para o enfrentamento da zika, dengue e chikungunya. A entrevista aconteceu em seu local de trabalho e durou cerca de uma hora.

P07 demonstrou ter conhecimento sobre as práticas da ciência aberta e ser favorável a elas. O pesquisador considera que a ciência aberta é uma tendência no ambiente acadêmico, iniciada pelos periódicos de acesso aberto. Essa foi uma nova categoria que se relaciona com os aspectos da cultura acadêmica. Apesar de considerar como algo positivo, P07 apresenta de forma sutil sua crítica a esse novo modelo de negócio dos periódicos, cujo acesso continua sendo pago. A responsabilidade do pagamento, que antes era do leitor, apenas foi transferida para o autor.

Eu acho que é uma tendência sem volta. Sem retorno. Eu acho que desde que começaram as revistas de acesso aberto. Agora a gente tem que pagar por elas, né? Para ter esse acesso ou pelo menos pra disponibilizar o paper. Então, na verdade, ao invés do leitor pagar para acessar a gente paga para publicar. Inversão.

A crítica apresentada pelo autor refere-se ao modelo de negócio via APAC, discutido nesta tese a partir do posicionamento de Babini (2014), Weitzel (2014) e Harvie et al. (2013). Em síntese, os autores apontam que tal prática representa um retrocesso para os objetivos do acesso aberto, pois é apenas uma nova forma de manter os antigos negócios lucrativos para as editoras comerciais.

Embora a crítica seja apresentada logo no início da entrevista, P07 indicou que a cultura, na qual se localiza os periódicos de acesso aberto, também promoveu o uso dos *pre-prints* dos artigos como forma de trazer a filosofia aberta para a cultura acadêmica. Além disso, o pesquisador aponta outro fator determinante para o contexto, que é a avaliação das agências de fomento que beneficiam pesquisadores que disponibilizam sua produção em repositórios de *pre-print*.

Mas hoje nós temos também revistas no sistema *pre-print*, que também antes de publicar os autores já disponibilizam os artigos. Isso vem crescendo também. Inclusive algumas agências de fomento, não sei se no Estados Unidos, eu não recordo agora, mas elas estão induzindo,

estão pontuando mais os pesquisadores que fazem essa disponibilização *preé-print*. Então eles ali, quando eles vão avaliar um projeto, quem tem artigos no sistema *pre-print* tem uma pontuação maior. É um diferencial. Então eu acho que é um caminho sem volta isso aí, né? E a pesquisa quanto mais aberta, melhor.

Na percepção de P07, a tendência à ciência aberta, inclusive para os dados, justificar-se-ia pelo aceleração na divulgação dos resultados da pesquisa e, conseqüentemente, do avanço da ciência.

Eu acho que disponibilizar os dados também torna a ciência mais fácil. Faz com que as coisas ocorram mais rápido. E o que que o pesquisador quer no final? Ele quer na verdade um feedback sobre aquele estudo, sobre aqueles dados que ele publicou, da maneira mais rápida.

Com isso, o pesquisador declara sua prática de disponibilizar seus resultados em sistemas de *pre-print*, bancos de dados, redes sociais e eventos científicos. Especialmente os dados são depositados à medida que vão sendo produzidos, ou seja, imediatamente.

P07: Se a gente tivesse que esperar eles concluírem o trabalho pra publicar, esse tempo seria muito maior e aqui essa divulgação eles tão fazendo. Não somente pelo site, eles tão fazendo essa divulgação por blog também e via Twitter [...] A gente sempre gostou de disponibilizar os dados, então, por exemplo, nesse, no meu antigo laboratório [nome instituição], eu sou um dos responsáveis, eu era pelo menos até um dos responsáveis por essa página, então a gente mantinha uma página, a qual era uma descrição da equipe a todo o material. [...]

MC: Vocês depositam na hora que vão publicar o artigo ou vão depositando a cada momento que vão sequenciando?

P07: Não, antes. A gente deposita a cada momento...

Outra vantagem apontada pelo pesquisador para as práticas foi o aumento da visibilidade e a rapidez no acesso às informações. A visibilidade traria como consequência desejável melhores métricas de avaliação dos pesquisadores, tais como taxa de citação. Percebe-se aqui a expectativa que tais valores sejam convertidos em recompensas de valor acadêmico para os pesquisadores.

A gente sempre disponibilizou essas publicações como uma maneira que... hoje tudo é medido por métricas, né? Olham quantos artigos você tem, qual índice de citação desses teus artigos. Eu acho que essa é uma iniciativa, eu acho que outros grupos que fazem isso também, isso já auxilia na divulgação, mas principalmente para qualquer pessoa possa rapidamente acessar aqui e ter um histórico do grupo de pesquisa e tal. Mas hoje a gente tem o ResearchGate também que é uma forma também de divulgação. Então isso acaba, acaba fazendo com que o trabalho que tu fazes se torne então mais visível.

Tal como P01, P07 apontou a idade dos pesquisadores e o uso de tecnologias de comunicação como um fator interveniente na comunicação aberta dos dados. No

entanto, ele acrescentou outros dois elementos novos que estão influenciando esse cenário: a atualização das políticas editoriais e o desenvolvimento de banco de dados.

A gente tem um banco de dados chamado de Ibank, que é um banco de dados emitido pelo governo americano. É uma estrutura onde estão todas os genomas de todas as espécies de micro-organismos, espécies maiores são depositadas. É um banco de acesso livre, então qualquer pessoa pode acessar lá e pegar esses dados. As revistas antes de você publicar um paper com fragmentos do genoma, elas te obrigam, entre aspas, que você faça o depósito dessa sequência para que de fato ela fique disponibilizada.

Apesar dos elementos favoráveis, P07 salientou que alguns pesquisadores são receosos com a comunicação aberta dos dados em razão da perda do ineditismo no uso dos dados. O pesquisador declarou já ter sido temeroso em relação a esse fator, mas considera que o receio foi deixando de existir à medida que ganhou experiência em publicar artigos.

Há um medo, assim, por parte de alguns pesquisadores, eu confesso que eu já tive esse medo, principalmente quando a gente tá na fase do mestrado e doutorado, né? Que você tem medo, “não, vão pegar meus dados, vão publicar meus dados”...

É possível perceber que a preocupação apontada por P07 está diretamente relacionada com as implicações da publicação de artigos de periódicos. O tema foi ressaltado pelos outros pesquisadores entrevistados, especialmente porque tais publicações são necessárias para a avaliação da carreira dos pesquisadores e são consideradas como meios efetivos para conferir reconhecimento e recompensas aos seus trabalhos.

Embora o receio seja consenso entre os pesquisadores entrevistados, P07 apresenta um paradoxo. A ampla divulgação dos dados os deixam mais suscetíveis de serem utilizados primeiramente por outros pesquisadores mal-intencionados. No entanto, no momento que disponibiliza os dados, o pesquisador está se registrando como seu legítimo produtor. Assim, seria mais fácil identificar os usos indesejados. Além disso, P07 afirma que o reuso dos dados demanda a contextualização da pesquisa, dependendo assim da comunicação direta com o pesquisador que coletou os dados.

Então eu acho que quanto mais cedo a gente publicar esses dados, disponibilizar isso aí, melhor. Há casos, por exemplo, de pessoas que podem ir lá e, entre aspas, “roubar os seus dados” e publicar? Pode. A gente está susceptível a isso, pessoas de mau caráter em qualquer área, inclusive na ciência. Mas essas pessoas podem ser facilmente desmascaradas. Não adianta eu pegar só esse dado, só o dado do sequenciamento se eu não tiver um conjunto de fatores. Ah, tudo bem, eu coloco aqui a origem desses animais e a data da coleta, mas e as características desses animais? Eram fêmeas? Eram machos? Qual que era a idade desses animais, em que

regiões especificamente eles foram coletados? Essas informações geralmente a gente disponibiliza só quando a gente publica o artigo. É claro que se outro pesquisador quiser esses dados a gente pode falar...

De forma geral, esses foram os elementos apontados por P07 em sua entrevista. O conjunto dos fatores foram sistematizados no Quadro 40.

Quadro 40: Codificação aberta da entrevista - P07

Dimensão	Categoria preliminar	Categoria aberta	Propriedade
Contexto social da doença	Especificidade do vírus da zika	Microcefalia, condições favoráveis e doença nova	
	Saúde pública	Emergência de saúde pública	Rapidez na comunicação
Imperativos da ciência	Aceleração da ciência	Aceleração da ciência	Finalidade da pesquisa, feedback Vantagem
	Aspectos éticos	Aspectos éticos	Maximização dos dados Exigência dos periódicos
		Idade dos pesquisadores Tendência de acesso aberto	Periódicos de acesso aberto e <i>pre-prints</i>
	Cultura acadêmica	Diferenças disciplinares (tipos de dados)	Processamento das amostras dos primatas Desvantagem (roubo de ideia)
		Ineditismo e parasitismo	Agências de fomento e os <i>pre-prints</i> (experiência EUA)
	Recompensas	Recompensa	
Reuso	Reuso	Colaboração	Contexto da pesquisa (colaboração direta) Comparação internacional
	Transparência para ciência	Utilidade dos dados	Comparação
		Divulgação	Identificação de erros (receio)
Infraestrutura	Sistema de informação	Artigos científicos	Política editorial obrigatória para a divulgação dos dados
	Gestão dos dados de pesquisa	Gestão dos dados de pesquisa	Profissionais, financiamento e padronização

Fonte: Elaboração própria.

5.1.7. Considerações acerca da entrevista com P08

A oitava pesquisadora entrevistada tem formação na área de medicina, com ênfase em saúde pública. Seu projeto de pesquisa abarca temas relacionados à epidemiologia e aos aspectos da saúde pública. Além de pesquisadora, ela é editora de

dois periódicos de acesso aberto em saúde pública, um periódico brasileiro e outro estrangeiro.

A entrevista com P08 ocorreu via Hangouts⁴⁷ e durou cerca de uma hora (Apêndice H). A pesquisadora mostrou-se familiarizada com o tema da ciência aberta e citou sua experiência como editora. Em decorrência da sua atuação, P08 mostrou-se favorável à prática de comunicação dos dados de pesquisa.

Eu sou totalmente a favor da abertura completa de todos os dados. Eu acho que isso é uma questão de princípio em defesa da ciência. Não é em defesa de mais nada, é em defesa da própria ciência.

A defesa da ciência foi um dos fatores apontados pela pesquisadora para justificar seu posicionamento favorável. Além desse, foi apontado que diversos periódicos internacionais têm requisitado os dados relacionados aos artigos que publicam. Foi dado como exemplo a prática dos periódicos PLOS.

Embora a pesquisadora tenha apresentado uma perspectiva amplamente favorável, foi apontado um conjunto de fatores que problematizam o cenário da ciência aberta e que, portanto, são empecilhos. Entre eles, destacam-se os seguintes:

- ausência de recompensa para os pesquisadores que coletaram os dados;
- presença dos pesquisadores parasitas;
- ausência de profissionais especializados para o tratamento dos dados que possam ser contratados pelos projetos de pesquisa;
- perda do ineditismo;
- identificação de erros e embasamento de críticas ao trabalho que apresentou os dados;
- falta de infraestrutura adequada.
- aplicabilidade dos dados em outras pesquisas e as diferenças disciplinares.

As problemáticas aparecem nos três macroprocessos da comunicação dos dados. Segundo P08, os dados devem receber um tratamento especial para possibilitar o seu reuso.

MC: ... é diferente o tratamento que você dá aos dados se você pensa em compartilhá-los? Você tem um esforço a mais pra preparar ele pra compartilhar?

P08: Claro, eu tenho que fazer um dicionário bom de variáveis, um dicionário de variáveis decente. Eu tenho que recategorizar variáveis. Porque quando a gente faz a pergunta, a gente se baseou muito nas grandes pesquisas nacionais, IBGE, FINAD, não sei o quê, para ter um

⁴⁷ O Hangouts é um software para realização de videochamadas.

questionário validado. Mas tem coisa que não funciona no Rio de Janeiro e tem coisas que a gente descobre a caminho e que a gente codifica a posteriori. Então tem todo um tratamento.

A forma como os dados serão distribuídos pode variar de acordo com o tamanho da pesquisa. P08 citou dois exemplos que ilustram essa diferença. Em uma pesquisa grande, os dados eram tratados por um centro de dados que, além desse processo, se responsabiliza por distribuir os dados sempre que considerasse o pedido pertinente.

O estudo [A] tem um centro de dados, que faz a gestão do dado. Faz tudo, desde a entrada do dado, até organizar o dado, limpar o dado, arrumar o dado. Faz toda aquela parte de pré-processamento insuportável e depois cede o dado pra uns. Nós temos um comitê. O [estudo A] tem um comitê de publicação que funciona como um filtro. As pessoas submetem um projeto, um artigo, um resumo, é uma coisa pequenininha, tem no sistema, nada além. Submete, aí isso é avaliado, como se fosse uma avaliação de pares, né? Alguém dá um palpite, outro dá outro. Ninguém dá palpite do conteúdo exatamente, entendeu? Mas dá palpite assim: “Olha, faz sentido, não faz sentido, nós temos dados pra isso ou não temos”. Dá uma organizada ou até dizer o seguinte: “Olha, o projeto eu tô dentro”. [...] e aí uma vez o projeto aprovado o centro de dados manda o dado para a pessoa. Encriptado. Tem as regras, né? Pra evitar confusão e proteger o dado mesmo, proteger o participante. Não vem com o nome, tudo bem. Mas mesmo assim tem que ter cuidados e critérios.

No exemplo do estudo A, existe um setor dentro da instituição da pesquisadora que se preocupa com o tratamento e os aspectos éticos necessários para distribuir os dados. Em contraste, o outro exemplo citado por P08 é de um estudo pequeno, que não conta com a mesma estrutura.

A outra pesquisa é uma pesquisa pequenininha que eu fiz, aí eu era coordenadora de um inquérito de saúde na comunidade [Y] em 2012, por aí, 2013. Quem organizou, limpou e arrumou o dado foi eu. Eu e as alunas. Eu ensinei a elas a limpar dados e organizar dados na pesquisa [...] Esse dado é completamente aberto, livre, quem quiser usar pode usar à vontade e tal. Mas é amador, não é profissional, entendeu? Então alguém tem que mandar o e-mail pedindo. Quem me pediu, recebe o dado inteirinho.

Embora P08 exerça práticas de produção e compartilhamento favoráveis à abertura dos dados, ela declarou não se beneficiar dos dados de outras pesquisas. De acordo com sua perspectiva, é muito trabalhoso utilizar dados de outras pesquisas, já que os dados precisam ser contextualizados e, normalmente, atendem somente às especificidades da pesquisa que os gerou. Como alternativa ao impasse, P08 afirma que, quando tem interesse em dados coletados por outros pesquisadores, ela propõe colaboração. Portanto, para P08 o reuso dos dados pressupõe colaboração direta entre os pesquisadores.

Eu não utilizo dados de outras pesquisas [...] Tenho trabalho pra cacete, Deus me livre. Porque você tem que entender o que era a outra pesquisa. Para que que eu vou fazer isso? Eu tenho tanta coisa aqui na minha vida. [...] Mas agora, por exemplo, se eu tô com uma pergunta de pesquisa que eu quero entender melhor, aí eu posso usar sim, posso até pedir. Se a gente tá querendo fazer um estudo, aí, de espaço e tempo, de clima, tranquilamente eu pego o telefone, ligo pra Pernambuco. Mas é pro amigo, né? Digo assim: "Olha, nós estamos querendo isso aí. Vocês têm também? Vamos fazer junto? Em geral é vamos fazer juntos? [...] Não é usar seus dados. É vamos fazer junto? Porque é ele quem entende o dado dele.

P08 pondera ainda que o reuso dos dados sem colaboração não é aplicável para o seu campo, a saúde pública. No entanto, pode ser aplicável em outras áreas, tais como biologia e estatística.

Eu acho que esse papo de "vou usar seu dado" também é muito estúpido. Porque você tem que entender aquilo lá, entendeu? Eu duvido que um chinês consiga pegar o estudo [A] e fazer qualquer trabalho que preste. Não vai. Porque não vai entender o contexto. [...] Na área biológica, não. Na área biológica ele consegue, é muito mais genérico.

Todas as categorias deduzidas da entrevista de P08 e que serão utilizadas no estudo estão sistematizadas no Quadro 41.

Quadro 41: Codificação aberta da entrevista - P08.

Dimensão	Categoria preliminar	Categoria	Propriedade
Contexto social da doença		Virologia brasileira	Bom exemplo de comunicação aberta dos dados sobre zika
		Poder de dispersão da doença	Especificidade da zika Espalhar pelo mundo
		Microcefalia	Comoção pública Análise do propósito da pesquisa que se beneficiará dos dados comunicados
		Função social	
Imperativos da ciência	Aspectos éticos	Aspectos éticos	Anonimização dos dados
		Ineditismo	Forma de recompensa
	Aceleração da ciência	Defesa da ciência	Princípio para ciência
	Recompensa	Recompensa acadêmica	Reconhecimento para quem coletou os dados
		Parasitas	Pesquisadores que não colaboraram com a pesquisa e se beneficiam dos dados Utilidade para outras pesquisas

Capítulo 5 - Fatores que influenciam a comunicação dos dados de pesquisa sobre o vírus da zika na perspectiva dos pesquisadores

		Parasitas internacionais	Pesquisadores estrangeiros que querem apenas os dados dos pesquisadores brasileiros
	Transparência	Identificação do erro	Crítica ao trabalho
	Cultura acadêmica	Diferenças disciplinares	Expansão da pesquisa e crítica
	Reuso	Reuso	Contexto da pesquisa e colaboração
	Tratamento dos dados	Demanda mais esforço se os dados forem comunicados abertamente	
	Políticas de informação	Políticas editoriais	Promovem a comunicação aberta dos dados
	Profissionais dos dados	Perfil dos profissionais	
Financiamento	Custos	Custos	Compartilhar os custos para compartilhar os dados
	Custos	Reconhecimento	Divisão entre quem coleta e quem analisa Forma de reconhecimento
Fenômeno		Profissionais para os dados	Restrição para contratação
		Armazenamento dos dados	Tratamento profissional
		Disponibilização dos dados	Condicional ao tamanho da pesquisa
		Contexto da pesquisa	Demanda colaboração

5.1.8. Considerações acerca da entrevista com P09

O nono pesquisador entrevistado também tem formação em medicina com ênfase em saúde pública. Seu projeto de pesquisa consiste na investigação da incidência do vírus da zika em comunidades do Distrito Federal. Segundo P09, a pesquisa é realizada em parceria com o governo local.

O pesquisador declarou não ter conhecimento sobre as práticas de comunicação aberta dos dados de pesquisas em seu campo de atuação.

MC: [...] Mas os dados brutos, o conjunto de dados, depositados em outro espaço, em um repositório ou uma central de banco de dados, enfim, na sua área, não é comum isso?

P09: Não, nunca vi isso aqui, nunca vi isso. [...] Em um projeto de pesquisa não, eu tenho que pegar dados de uma pessoa e ela precisa e aí ela precisa assinar, que permite que eu colha sangue, que eu armazene sangue. Agora, o que você tá querendo nós não sabemos como fazer. Nem imaginamos o que precisa. Porque ninguém aceita na minha área, eu nunca ouvi isso: “olha, os dados tão aqui, faz o que você quiser com eles”. Não existe.

A prática de comunicação dos dados conhecida pelo pesquisador se dá por meio de artigos de periódicos, comunicação em eventos, teses e palestras. Com isso, conclui-se que não existe divulgação dos dados brutos das pesquisas. A prática está de acordo com as exigências da sua área de conhecimento que demanda que os dados sejam tratados e analisados. P09 aponta ainda que não existem sistemas de informação adequados apenas para os dados de pesquisa. Portanto, somente os dados mais importantes são divulgados no artigo ou documento científico.

P09: Os dados brutos são trabalhados [...] então, a gente vê o dado, colhe e tal, e coloca no artigo. A gente não coloca esses dados. Ele fica no nosso computador e também no dele, né? Do outro autor. Os dados brutos.

MC: Mas eles não são divulgados em outros espaços

P09: Não, não tem outro espaço pra isso, não existe. Quer dizer, as revistas só aceitam é o artigo, né? Sob forma de artigo.

O pesquisador P09 pontuou que não é de interesse dos pesquisadores apresentarem apenas os dados em um sistema de informação que não seja de publicações científicas tradicionais, uma vez que os pesquisadores seriam avaliados academicamente apenas por essas publicações e não pelo dado bruto. No entanto, os dados brutos são os insumos necessários para a publicação. Com isso, é necessário manter os dados inéditos para que seja possível produzir publicações que gerem recompensas aos pesquisadores.

P09: [...] Quem produziu o dado quer trabalhar o dado, porque hoje nós somos vigiados pelas publicações. E quem não publica se trumbica. Tá certo? Então você não pode nem ser orientador senão tiver publicação.

MC: E o senhor acha que se abrir esses dados você perderia em publicação?

P09: É claro, porque outras pessoas vão usar seus dados. Quer dizer, eu acho que quem produz o dado quer usar o dado. E a originalidade uma das características do artigo também. Eu nunca vi isso na minha área. Alguém falar: “olha, eu tenho um banco de dados aqui”, a não ser os bancos de dados oficiais, dos sistemas de informações do ministério que são divulgados.

Os bancos de dados oficiais citados por P09 são sistemas mantidos e divulgados pelo Ministério da Saúde do Brasil, que apresentam dados sobre doenças com grande incidência no país.

Todas as categorias deduzidas da entrevista de P09, e que serão utilizadas no estudo, estão sistematizadas no Quadro 42.

Quadro 42: Codificação aberta da entrevista - P09.

Dimensão	Categoria preliminar	Categoria	Propriedade
Contexto social da doença		Expansão do vírus	Possibilidade de chegar em qualquer país
		Microcefalia	Aborto e diminuição de nascimentos
Financiamento		Interesse dos EUA	Ocorrência de zika na Flórida Investimento norte-americano
Imperativos da ciência	Aspectos éticos	Aspectos éticos	Privacidade dos participantes – termo
	Cultura acadêmica	Diferenças disciplinares	Inexistência de comunicação aberta na área de área de doenças infecciosas Forma de reconhecimento pela pesquisa Necessário para produção de artigos Única forma de avaliação da produção
	Reuso	Metanálise	Reuso dos dados
Infraestrutura	Sistema de informação	Banco de dados do governo	Sistema de informação sobre doenças e estruturas de saúde pública
Fenômeno		Artigos científicos	Forma de divulgar os dados de pesquisa na saúde pública (epidemiologia) Dados brutos precisam ser trabalhados Ausência de sistemas de informação para os dados
		Armazenamento dos dados	No próprio computador

5.1.9. Considerações acerca da entrevista com P10

A última entrevista foi realizada com o pesquisador P10, que tem formação em farmácia e biotecnologia. Seu projeto de pesquisa tem por objetivo o desenvolvimento de vacina contra o vírus da zika. É importante ressaltar que a pesquisa de P10 não é a mesma dos pesquisadores P03 e P04. Portanto, este estudo entrevistou pesquisadores de dois grupos diferentes que possuem o mesmo objetivo: o desenvolvimento de vacina contra o vírus.

O pesquisador P10 declarou ter entrado em contato com o tema da ciência aberta recentemente, por meio de uma palestra realizada em sua instituição. Ele acredita que a abordagem está sendo crescente em seu campo de atuação, especialmente para os casos de emergência de saúde pública. No entanto, o pesquisador ponderou que o foco da sua pesquisa difere da grande parte das outras

pesquisas sobre o tema. Isso porque ela é focada no desenvolvimento de tecnologia para a vacina, o que a faz ser essencialmente mais restrita.

[...] Tem gente que trabalha com essa área, trabalha com amostras clínicas e tudo, mas o nosso enfoque é o desenvolvimento de tecnologia. É o desenvolvimento de soro e vacina contra o zika. Já nessa área, a coisa não é tão *open science* assim, porque, quando você faz pesquisa de tecnologia você tem que pensar em proteção intelectual para aquilo virar um produto. Porque sem a proteção intelectual o produto perde a lucratividade.

As restrições de acesso aos resultados das pesquisas seriam provocadas pela expectativa de desenvolvimento de produto lucrativo. Portanto, os resultados das pesquisas precisam estar assegurados pelos direitos de propriedade intelectual, que garantem exclusividade na exploração da ideia. Segundo P10, essa é a prática recorrente de produção de tecnologias para saúde. No entanto, o caso específico da vacina contra o vírus da zika tem provocado outra forma de colaboração científica, na qual determinados grupos trocam experiências e informações diretamente. Nesses casos, as informações, bem como os dados não são abertos a qualquer interessado, apenas a um determinado grupo e de forma seletiva.

[...] uma iniciativa de um órgão americano que se chama Barda⁴⁸, que é um órgão que é ligado com essas questões de epidemiologia. Ele é um superior ao Medicine Institute of House, nos Estados Unidos, e esse pessoal, eles estiveram aqui no Brasil. A gente teve uma conversa com eles e a gente fechou um acordo de cooperação, no qual eles se comprometiam a compartilhar informações não cruciais, né, e não confidenciais, e outros produtores desenvolvedores de vacina com a gente e vice-versa. Então constantemente eles nos perguntam coisas sobre o nosso desenvolvimento e nos dão indicativos de como desenvolver naqueles pontos em que a gente apresenta dificuldade, né?

MC: Para vacina específica da zika?

P10: Para vacina da zika, isso só acontece pro zika. Não acontece pra outras coisas por enquanto. Mas é uma experiência interessante, quer dizer, é uma coisa que a gente tem visto que pode elevar o desenvolvimento mais rápido dessa parte tecnológica em si.

Um dos motivos apontados por P10 para essa nova dinâmica na comunicação científica é decorrente da classificação do tema como uma emergência de saúde pública. Tal motivo teria provocado o pesquisador e seu grupo a exercerem outra forma de colaboração científica, bem como tem alterado a forma de comunicação tradicional dos resultados das pesquisas para o grande público. O pesquisador afirmou que até as

⁴⁸ A missão da instituição é facilitar a comunicação entre as agências do governo federal e as partes interessadas do público para superar as questões de emergência de saúde pública no país. Fonte: <https://www.medicalcountermeasures.gov/barda.aspx>

revistas científicas mais tradicionais estavam promovendo rapidamente a publicação dos artigos e dos dados relacionados aos artigos que tratavam sobre o tema.

Agora com a história do zika, a gente realmente viu uma progressão nesse sentido. Vários trabalhos que foram publicados ao mesmo tempo em que foram submetidos. Então dava pra ver que estava submetido, mas já tava disponível. Os dados estavam disponíveis pras pessoas já adiantarem as suas próprias pesquisas. O que é uma coisa de certa forma importante. Porque se você for pensar bem, quanto tempo depois da pesquisa realizada ela é publicada de fato, muitas vezes isso tem uma demora aí que é postergada. Então quando você tem um problema de emergência médica, que é o zika, isso pode fazer a diferença.

O que tem motivado essa comunicação mais rápida, aberta e colaborativa é a expectativa do aceleração na produção de uma tecnologia capaz de superar os problemas causados pelo vírus. A excepcionalidade do vírus da zika em relação a outras doenças, também classificadas como emergências de saúde pública, seria decorrente da alta capacidade de dispersão do vírus pelo mundo, especialmente por transmissão sexual. A possibilidade de o vírus atingir grande parte da população, inclusive dos países desenvolvidos, teria despertado o interesse da indústria farmacêutica para a produção de novas tecnologias e fármacos.

As preocupações em relação ao vírus atingiram também os pesquisadores dos países desenvolvidos, com isso as práticas de colaboração estariam mais presentes. No entanto, P10 alerta que o interesse dos pesquisadores estrangeiros pelo seu trabalho está relacionado com o acesso aos dados coletados aqui.

A gente tem recebido muita gente de fora, pesquisando zika, chikungunya. O pessoal vem aqui e aí eles falam, "o que que vocês tão fazendo?" A gente tá fazendo isso, fazendo aquilo. E eles, "aham, aham, aham". Até o momento que você fala assim: "olha, e aí a gente coletou umas amostras de paciente soro lá, não sei aonde". E aí eles falam: "opa". Quer dizer, é isso que eles querem, né? Eles querem a amostra, eles querem o dado. Eles querem essas coisas, porque o resto eles já têm.

O pesquisador P10 ainda sinaliza que esses dados são disponibilizados para os pesquisadores estrangeiros, porque o Brasil não conta com os mesmos recursos necessários para analisar o dado com tanta rapidez quanto nos países desenvolvidos.

A gente tem um problema, a gente tem as amostras das coisas e não tem o que eles têm ou tem limitadamente. E então a gente precisa abrir pra fazer essa troca. [...] Esse é o principal. A gente não tem a capacidade de análise, isso é muito importante. E uma outra coisa que é muito importante também, a gente tem o acesso ruim a reagentes. Mesmo aos equipamentos, é ruim, demora muito. Os fornecedores brasileiros são péssimos, o mercado brasileiro é ruim, então eles não têm estoque local ou cada vez que você precisa de um reagente é sessenta dias. Lá não dá uma semana, pensa na diferença. Então é muito diferente e às vezes chega aqui e você não

sabe por que não funciona, não funciona, não funciona e você vai lá fazer a pesquisa como muitas vezes eu fui e disse: “poxa, faço a mesma coisa na França que eu faço no Brasil e lá dá certo e aqui não dá, por quê?” Porque o reagente que chega aqui, chega bichado, né? Desembarca lá não sei quando e vai e fica não sei quantos dias no aeroporto pra Anvisa liberar. Quer dizer, isso aí vai acabando com o sistema. Fora que pra gente custa cinco vezes mais que pra eles. Além de eles terem mais dinheiro, pra gente custa mais caro.

Percebe-se com isso que, nesse caso, a colaboração internacional é necessária, mas não é necessariamente interessante para os pesquisadores brasileiros. Assim como os outros pesquisadores entrevistados, P10 sinalizou que a expectativa de disponibilização dos dados coletados demanda mais esforços no registro dos dados. Os registros dos dados são feitos de duas formas: imagens registradas no computador e dados anotados em um caderno denominado “livro de registros”. Segundo P10, apenas alguns desses dados anotados à mão são registrados posteriormente no computador, mas, para que os dados sejam disponibilizados, todos eles deveriam ser digitalizados e descritos de forma padronizada.

Corroborando as informações apresentadas por P03 e P04, o principal canal para comunicação dos resultados, incluindo os dados de pesquisa, são os artigos de periódicos. Assim, a área da farmácia, especialmente aquela que desenvolve tecnologias para fármacos, não tem por cultura disponibilizar os dados brutos das pesquisas.

Todas as categorias deduzidas da entrevista de P09 e que serão utilizadas no estudo estão sistematizadas no Quadro 43.

Quadro 43: Codificação aberta da entrevista - P10.

Dimensão	Categoria preliminar	Categoria	Propriedade
Contexto social da doença	Saúde pública	Emergência pública Vacina e tecnologia contra o vírus	Provoca discussão sobre <i>open science</i> Demanda proteção intelectual para a produção do produto lucrativo Transmissão sexual Período de risco Proliferação além dos tópicos (principal motivo de interesse estrangeiro) Interesse da indústria farmacêutica
		Alta dispersão do vírus	
Imperativos da ciência	Reuso	Colaboração internacional	Interesse de outros países, especialmente dos EUA, pelo desenvolvimento da vacina
	Cultura acadêmica	Interesses comerciais	Vacina como produto lucrativo
	Aceleração da ciência	Aceleração dos resultados	Colaboração

	Recompensas Aceleração da ciência Aspecto ético	Ineditismo Avanço da ciência Aspecto ético	Condição para publicar o artigo Apresentação dos erros Os dados são anônimos
Infraestrutura	Sistema de informação Sistema de informação Sistema de informação Sistema de informação	Artigo Patente Publicações científicas e zika Banco de dados	Canal de comunicação dos resultados e dos dados Condição para divulgar os dados Artigos publicados imediatamente e de acesso aberto sobre a zika Restrição nacional
Financiamento	Custo	Divisão internacional da ciência Capacidade de análise dos dados	Interesse dos estrangeiros apenas nos dados Diferenças regionais

Fonte: Elaboração própria.

5.1.10.A percepção de pesquisadores a respeito dos fatores que influenciam a comunicação dos dados de pesquisa

A codificação aberta dos dados resultou em um conjunto de 29 categorias, das quais apenas sete foram consideradas novas, ou seja, não foram contempladas nas categorias preliminares (Quadro 34). Quatro das sete categorias emergentes dizem respeito ao contexto social da doença (Quadro 44 – laranja), tema objeto deste estudo. A singularidade do contexto aqui investigado justifica, portanto, o aparecimento de novas categorias, uma vez que nenhum dos artigos analisados na revisão sistematizada da literatura tratou do mesmo tema objeto de pesquisa.

As outras três categorias emergentes destacaram fatores influenciadores da comunicação dos dados de pesquisa que não se restringem ao contexto das doenças emergentes (Quadro 44 – azul). Uma delas diz respeito ao idioma pelo qual os dados devem ser registrados, caso haja a expectativa de comunicá-los abertamente. A imposição de um idioma apresenta-se como um empecilho específico para os pesquisadores que não o tem como língua oficial. As outras duas categorias tratam de pressupostos para obtenção de reconhecimentos e recompensas pelos esforços da pesquisa, pontuados pelos entrevistados como sigilo e ineditismo dos resultados.

Além das sete novas categorias, outras três tiveram sua relevância para a discussão do tema significativamente ampliada, embora tenham sido evidenciadas de forma preliminar (Quadro 44 – roxo). Uma delas trata das condições de análise e coleta

dos dados que foram discutidas nas categorias preliminares de forma periférica. Entretanto, nos dados coletados por essa pesquisa, o aspecto tornou-se interveniente do processo, assumindo assim um espaço de centralidade na discussão proposta. As outras tratam dos reconhecimentos e recompensas almejados pelos pesquisadores e da relevância que as publicações dos resultados em periódicos acadêmicos assumem nesse processo.

Quadro 44: Categorias abertas.

Categoria preliminar	Categoria aberta
Aceleração da ciência	Aceleração da ciência
Custo	Análise dos dados
Aspecto ético	Aspecto ético
Aspectos legais	Aspectos legais
Sistema de informação	Banco de dados
Reuso	Colaboração (reuso)
Cultura acadêmica	Diferenças disciplinares
--	Doenças emergentes
Instituição financiadora	Financiamento
Saúde pública	Função social
Cultura acadêmica	Idade dos pesquisadores
Transparência da ciência	Identificação de erros
--	Idioma
--	Ineditismo
--	Interesse dos EUA
Aspectos legais	Interesses comerciais
Reuso	Metanálise (reuso)
--	Microcefalia
Reuso	Parasitas (reuso)
--	Poder de dispersão do vírus
Gestão dos dados de pesquisa	Profissionais
Recompensa aos pesquisadores	Publicações em periódicos
Recompensa aos pesquisadores	Reconhecimento e recompensa
Custo	Recursos
Reuso	Reuso
Serviço de informação	Serviço de informação
--	Sigilo da pesquisa
Sistema de informação	Sistema de informação
Padrão	Tratamento dos dados

Fonte: Elaboração própria.

As 29 categorias abertas representam os fatores que influenciam a comunicação de dados de pesquisa, segundo a percepção dos pesquisadores. Tais fatores representam os resultados do segundo objetivo específico do estudo.

5.2. Ordenação conceitual e codificação axial

As categorias abertas evidenciadas pelas categorias preliminares foram essencialmente analisadas a partir dos dados coletados. Portanto, a discussão preliminar serviu apenas de ponto de partida para a análise, não se findando nela mesma. Com isso, as categorias preliminares e abertas representam etapas do processo de análise dos resultados. A próxima etapa prevista pela metodologia da teoria fundamentada, proposta por Strauss e Corbin (2008), é a reorganização das categorias abertas por meio da codificação axial.

A codificação axial é o processo de relacionar as categorias às suas subcategorias, com o objetivo de reagrupar os dados que foram divididos no processo da codificação aberta. A reagrupação se dá em torno do eixo de uma categoria, que associa outras categorias hierarquicamente dependentes dela por meio de suas propriedades e dimensões. Nesse processo, as categorias que se tornam subcategorias têm por finalidade gerar explicações mais precisas sobre o fenômeno estudado (STRAUSS; CORBIN, 2008). Após o processo da codificação axial, as categorias abertas foram reagrupadas em sete categorias e nove subcategorias (Quadro 45). Todas elas serão discutidas a seguir.

Capítulo 5 - Fatores que influenciam a comunicação dos dados de pesquisa sobre o vírus da zika na perspectiva dos pesquisadores

Quadro 45: Categorias axiais.

Categoria aberta	Categoria Axial	Subcategoria axial
Aspecto ético	1. Colaboração	1. Aspecto ético
Aspectos legais		2. Aspecto legal
Análise dos dados		3. Condições de análise
Colaboração (reuso)		
Metanálise (reuso)		
Parasitas (reuso)		
Reuso		
Aceleração da ciência	2. Cultura acadêmica	
Função social		
Identificação de erros		
Diferenças disciplinares		4. Diferenças disciplinares
Idade dos pesquisadores		5. Idade dos pesquisadores e uso de tecnologias
Doenças emergentes	3. Doenças emergentes	
Microcefalia		6. Microcefalia
Poder de dispersão do vírus		7. Poder de dispersão
Financiamento	4. Financiamento	
Interesse dos EUA		
Interesses comerciais		
Profissionais		
Recursos		
Ineditismo	5. Reconhecimento e recompensa	8. Sigilo da pesquisa
Publicações em periódicos		9. Publicações acadêmicas
Reconhecimento e recompensa		
Sigilo da pesquisa		Sigilo da pesquisa
Banco de dados	6. Repositório de dados	
Serviço de informação		
Sistema de informação		
Idioma	7. Tratamento dos dados	
Tratamento dos dados		

Fonte: Elaboração própria

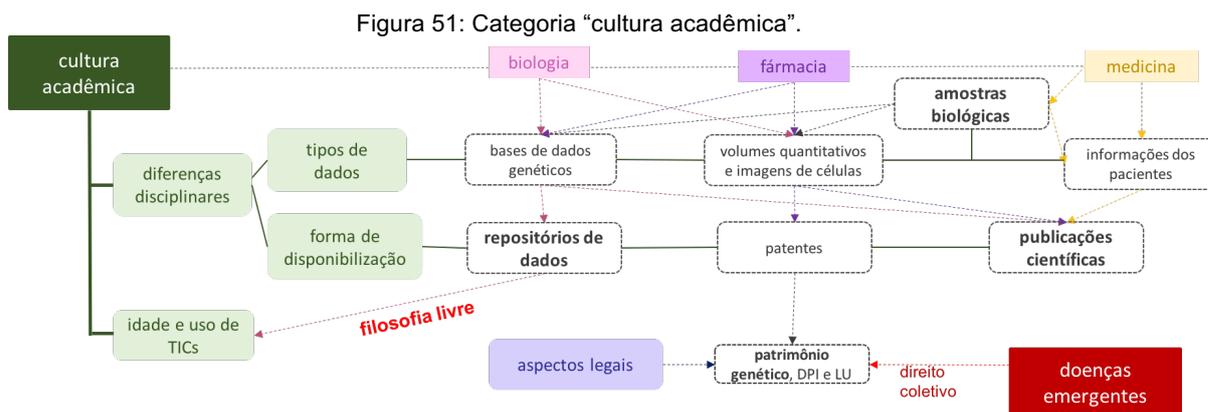
A reorganização das categorias abertas nas categorias axiais permitiu identificar os fatores convergentes e divergentes sobre a comunicação dos dados de pesquisa, na perspectiva dos pesquisadores especialistas sobre o vírus da zika. Tais fatores representam os resultados do terceiro objetivo específico do estudo. Em síntese, os fatores convergentes preponderam nas seguintes categorias: colaboração, financiamento, reconhecimento e recompensa e doenças emergentes. Os fatores divergentes evidenciaram-se especialmente em função das diferenças disciplinares e predominam nas seguintes categorias: cultura acadêmica, tratamento dos dados e repositórios de dados de pesquisa.

5.2.1. Categoria “cultura acadêmica”

A cultura acadêmica diz respeito aos elementos que afetam o comportamento dos pesquisadores, conforme definiu Dallmeier-Tiessen et al. (2014). A sistematização da categoria preliminar sobre o tema sugeriu quatro vertentes de expressão da categoria: as diferenças entre as áreas do conhecimento, o uso de tecnologias da informação e a idade dos pesquisadores e os interesses comerciais. Com isso os elementos que afetam o comportamento dos pesquisadores são pensados em função das práticas e dos interesses das áreas do conhecimento, bem como da capacidade no uso das novas tecnologias de informação e comunicação.

A análise dos dados da entrevista confirmou sua relação com as três primeiras vertentes. A discussão em torno dos interesses comerciais foi realocada para a subcategoria “financiamento”, subordinada à categoria “recursos”, que será apresentada ainda neste capítulo. As diferenças disciplinares foram discutidas a partir de dois elementos, a tipologia dos dados e as predominantes formas de comunicação dos resultados das pesquisas, com ênfase para os dados. Já a idade dos pesquisadores e o uso das tecnologias foram diretamente relacionados, uma vez que as perspectivas dos entrevistados sobre o tema evidenciaram a relação.

Os dados resultantes da pesquisa apontam que a área do conhecimento e a idade dos pesquisadores influenciam a abertura ou não dos dados de pesquisa. As diferenças entre as áreas do conhecimento consideradas para essa discussão dizem respeito aos tipos de dados necessários para as pesquisas e as formas aplicadas na divulgação dos resultados. O outro elemento que compõe esse fator evidencia que a utilização de repositórios de dados como forma de disponibilização dos dados favorece sua comunicação aberta. Os pesquisadores usuários de tais sistemas relacionaram a tendência ao uso dos sistemas por pesquisadores mais jovens, com mais familiarização no uso das TICs e das filosofias abertas (Figura 51).



Fonte: Elaboração própria.

A categoria cultura acadêmica é explicada por meio de suas propriedades levantadas preliminarmente no capítulo anterior. No contexto das pesquisas relacionadas ao vírus da zika a categoria relaciona-se ainda a duas subcategorias, que serão exploradas a seguir.

5.2.1.1. Subcategoria “diferenças disciplinares”

As diferenças culturais entre as áreas do conhecimento se dão especialmente por causa da natureza dos dados e da forma como os pesquisadores comunicam seus resultados de pesquisa. Entre os pesquisadores entrevistados, reuniram-se três áreas do conhecimento: biologia, medicina e farmácia.

Os pesquisadores P01, P02, P05⁴⁹, P06 e P07 estão essencialmente relacionados ao campo da biologia (grupo A). Em comparação às outras áreas aqui tratadas, o grupo apresentou maior propensão às práticas de comunicação aberta dos dados de pesquisa. Os dados oriundos dessas pesquisas estão relacionados majoritariamente à genética molecular, que resultam na sistematização das bases de dados. Mesmo nos casos em que são coletadas informações dos pacientes, os dados de pesquisa resultante são de genética molecular. Tais dados apresentam as seguintes características que os diferenciam dos demais: não identificam os pacientes; são disponibilizados sem a necessidade de análise; são aplicáveis para outras pesquisas. Geralmente, os dados são disponibilizados em repositórios de dados, tais como os bancos de dados genéticos. Conforme o relato histórico apresentado na revisão de

⁴⁹ O pesquisador P05 tem formação na área de medicina, mas sua atuação é diretamente relacionada ao campo da biologia.

literatura deste trabalho, os repositórios de dados genéticos foram uma das primeiras iniciativas internacionais que mobilizaram pesquisadores do mundo inteiro. Sua gênese, bem como seus ideais motivadores, contribui para o entendimento dos fatores que colocam o Grupo A mais próximo das práticas da comunicação aberta dos dados.

Os pesquisadores P08 e P09 atuam com pesquisas no campo da medicina (grupo B). Suas pesquisas dependem diretamente das informações dos pacientes. Em decorrência da sua tipologia, os dados podem demandar ou resultar na identificação dos pacientes e, frequentemente, precisam ser contextualizados para que sejam realizadas novas análises. Os dados são disseminados, necessariamente, em artigos de periódicos ou outras publicações acadêmicas tradicionais, que demandam sua prévia seleção. Portanto, para o campo da medicina não existem sistemas de informação para a disponibilização dos dados brutos da pesquisa.

Os pesquisadores P03, P04 e P10 realizam pesquisas relacionadas à área de farmácia, com o objetivo de desenvolver vacina para o vírus da zika (grupo C). Os dados necessários ao desenvolvimento das pesquisas abrangem aqueles descritos para o grupo A e os dados resultantes de experiências com substâncias. Em ambos os casos, os dados diferem daqueles do grupo B por não tratarem diretamente de informações dos pacientes e por não demandarem contextualização para análise. As duas características aproximam-se das especificidades dos dados do grupo A. No entanto, diferem-se pela natureza dos resultados que geram ou almejam alcançar. Os resultados têm por objetivo gerar produto patenteável e/ou comercializável, portanto requerem mais sigilo dos processos e dos resultados das pesquisas. Todos os macroprocessos da comunicação dos dados são afetados pelos objetivos específicos das pesquisas realizadas pelo grupo C. Conforme salientaram Childs e Mcleod (2004), os direitos de propriedade intelectual, por meio das patentes, regulam tanto as práticas de armazenamento, como de acesso e de reuso dos dados protegidos.

De forma geral, identificou-se que os tipos dos dados envolvidos nas pesquisas sobre o tema são de natureza quantitativa, bases de dados genéticos e imagens. Tais resultados aproximam-se daqueles apresentados por Whitmire, Boock e Sutton (2015), que, além dos já citados, acrescentaram os textos qualitativos para a área da saúde pública. Os canais de divulgação dos dados variam em função da natureza dos dados e da área do conhecimento, repositório de dados para o grupo A e patentes para o grupo C. Apesar das diferenças, observou-se que em algum momento todos os pesquisadores almejam publicar seus resultados em periódicos científicos.

5.2.1.2. Subcategoria “idade dos pesquisadores e uso das tecnologias”

A idade dos pesquisadores somada à suas habilidades para o uso de tecnologias de informação foi outro aspecto classificado como interveniente na perspectiva de dois pesquisadores entrevistados, P01 e P07. A relação traçada por P01 pressupõe que os pesquisadores mais velhos exercem cargos de chefia em seus laboratórios e se interessam por manter a exclusividade na exploração dos dados com vistas a sua publicação de forma inédita. Já os pesquisadores mais jovens, mesmo apresentando a mesma finalidade de publicação dos resultados, considerariam que a colaboração de outros pesquisadores e de seus dados faz com que seus resultados sejam melhores e, por isso, teriam mais chances de publicar.

P01: Eu acho que tem a ver um pouco com o interesse internacional e isso que eu te falei de ter esses pesquisadores um pouco mais jovens. Porque esses pesquisadores, que eu conheço, a maioria deles tá na faixa dos trinta e cinco, quarenta e cinco anos. Então, se você vai para um laboratório que o chefe de pesquisa ele tem uns cinquenta e poucos, sessenta, provavelmente ele é do tipo que não compartilha informação porque ele se mantém num nicho, porque se ele tiver essa informação, ele vai publicar antes ou ele vai conseguir publicar melhor. Mas o que eu acho é que esses pesquisadores mais jovens, o que eles estão tentando fazer é exatamente a mesma coisa, só que se a gente publicar junto a gente vai conseguir publicar melhor, ao invés de tentar um esforço isolado, eu me junto com alguns grupos e a gente publica isso...

O pesquisador P07 considerou que os pesquisadores mais jovens estão mais adaptados ao uso das novas tecnologias da informação e, conseqüentemente, às práticas colaborativas da ciência. O pesquisador considera que o ambiente da Internet promove uma nova cultura, na qual a globalização da ciência e o compartilhamento de informações são emergentes. Sua perspectiva relaciona-se com os fundamentos da filosofia aberta, discutida por Costa (2006). Tal filosofia abrange as práticas de comunicação científica, bem como de produção de software e compartilhamento de arquivos e informações no contexto da Internet.

P07: [...] as coisas estão cada vez mais globalizadas, a ciência sempre foi globalizada. O pessoal que é de uma geração anterior, que não tinha acesso à Internet ou que dependia de revistas físicas para ter acesso aos dados, tiveram um momento muito mais difícil e desde o advento da Internet e desses portais que surgiram aí, por exemplo como o Scopus e da Elsevier, que é a maior, a maior editora mundial fundada lá em mil quinhentos e pouco, então essas conglomerações elas têm muita força na divulgação dos dados. Então, inicialmente eles organizaram essas revistas e comercializavam os papers e com a questão agora dos papers abertos, aí eles tão tendo também que adaptar a essa realidade.

5.2.2. Categoria “colaboração”

As práticas de colaboração variaram significativamente entre as áreas do conhecimento. A variação ocorreu em função dos tipos de dados, da forma como eles são comunicados e dos interesses que motivam a pesquisa. De forma geral, a finalidade da colaboração é a produção de novos resultados de forma mais rápida e com eficiência no uso dos recursos.

O compartilhamento de recursos é um dos elementos que motivam as práticas de colaboração. A pesquisadora P04 explicou que, por causa da falta de recursos necessários para o desenvolvimento de algumas etapas da pesquisa, é oportuna a associação com outros pesquisadores que disponham da infraestrutura almejada. A associação implica o compartilhamento dos dados.

A associação é vista por P05 como algo inerente à própria ciência. Em sua perspectiva, a ciência, assim como os resultados das pesquisas, são universais e devem, portanto, circular livremente. O pesquisador também declarou ter facilidade em associar-se com grupos de outros países que dispõem dos recursos necessários, quando os seus são insuficientes.

P05: [...] eu tenho financiamentos do ministério, CNPq, mas são financiamentos pequenos, que dá pra gente fazer uma pesquisa. Eu não reclamo. Sempre tive financiamento, durante esses quarenta anos nunca deixei de ter dinheiro no laboratório, mas é porque eu não faço uma pesquisa cara. Eu não tenho grandes voos. Quando eu preciso de alguma coisa, eu me associo com um colega. Eu tenho estudantes em diversas áreas e fora [do país], então a gente faz associação. Eu mando fazer lá e a gente troca informação, dentro dessa história que eu tô te falando, eu acho que a pesquisa é universal.

Apesar da naturalidade com que são descritas as práticas de colaboração, quando elas acontecem com vistas ao compartilhamento de recursos, nota-se um elemento perverso de desigualdade social. Tais associações acontecem em decorrência da falta de recursos dos países pobres, subdesenvolvidos ou em desenvolvimento, e resultam na entrega dos dados e das informações para os países desenvolvidos. A disparidade da relação reside na determinação com que os países ricos disponibilizam equipamentos e os países pobres disponibilizam os seus dados e seus resultados de pesquisa.

P07: O que que eles têm de diferencial? Eles têm uma maior capacidade na geração desses dados, porque eles têm mais equipamentos, eles têm mais dinheiro.

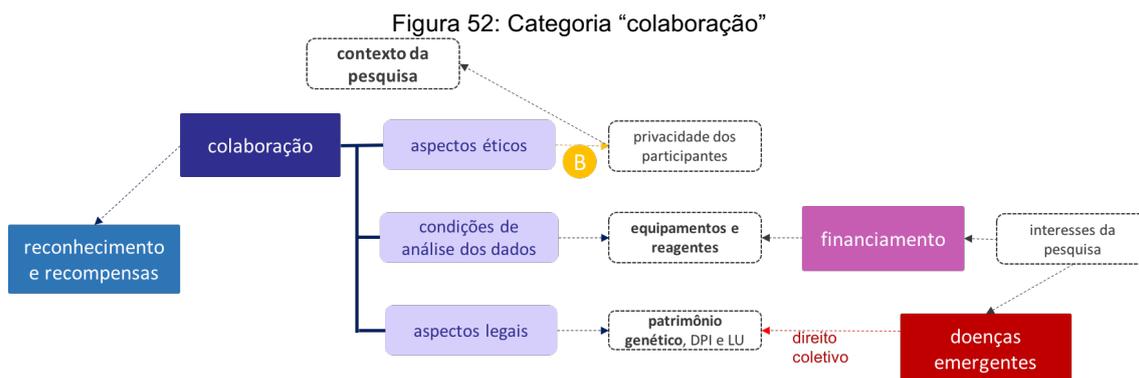
Outra motivação apresentada pelos pesquisadores entrevistados para a prática de colaboração é a necessidade de contextualização dos dados da pesquisa. Especialmente em áreas como a medicina, a utilização dos dados demanda maior detalhamento do contexto da coleta e dos atores envolvidos. Com isso, é mais estratégico convidar o pesquisador que coletou os dados para desenvolver o novo estudo em conjunto do que apenas utilizar os dados que ele disponibilizou.

P08: Não é usar seus dados, é vamos fazer junto? Porque é ele quem entende o dado dele. Eu acho que esse papo de “vou usar o dado” também é muito estúpido. Porque você tem que entender aquilo lá, entendeu? Eu duvido que um chinês consiga pegar o estudo [A] e fazer qualquer trabalho que preste. Não vai. Porque não vai entender o contexto. O que que ele vai entender que são quinze mil funcionários públicos de instituições federais, estaduais. O que o cara vai entender disso? Ele vai conseguir fazer algo que preste? Na área biológica, não. Na área biológica ele consegue, é muito mais genérico. Sobre epidemiologia.

Na fala da pesquisadora, é enfatizada a importância do contexto para um bom uso dos dados. Entretanto, ela considera que tal necessidade é específica do seu campo de atuação, a medicina, e que outras áreas, como a biologia, não teriam o mesmo imperativo. De fato, foi possível observar a diferença entre as duas áreas, especialmente em relação aos dados genéticos.

P07: A gente coleta amostras de primatas, não humanos, que no caso é o Projeto Zika que a gente tem aqui. [...] Então a gente tem pego todos esses animais que chegam doentes lá ou os animais que são capturados. A gente pega amostras desses animais e processa e aí a gente tem encontrado o vírus. É um dado não publicado ainda. Nesse momento, o que a gente tá fazendo agora a gente tá sequenciando, porque só a identificação molecular, quer dizer, só você encontrar o genoma ali pode ser uma contaminação, pode ser [outra coisa]. Então a gente precisa sequenciar e aí sim, a partir desses, com a sequência do genoma aí a gente disponibiliza essa sequência, esse é um dado bruto interessante porque daí qualquer pessoa em outra região do Brasil ou do mundo pode acessar esse genoma, pegar esse genoma e comparar com genomas que estão sendo estudados.

Portanto, a colaboração está implicada em três fatores. O primeiro é a necessidade de recursos para a análise dos dados de forma mais eficiente. Esse fator se relaciona com o segundo, referente aos aspectos legais do desenvolvimento da pesquisa e aos direitos de propriedade intelectual. O terceiro diz respeito à necessidade de entendimento do contexto da pesquisa, que, especialmente no campo da medicina, afeta diretamente a privacidade dos participantes, discutida a partir dos aspectos éticos da pesquisa (Figura 52).



Fonte: Elaboração própria

Todos os desdobramentos e propriedades relacionados à categoria colaboração serão evidenciados a partir de suas três subcategorias, descritas a seguir.

5.2.2.1. Subcategoria “aspectos da ética na pesquisa”

Um dos elementos sobre a ética na pesquisa que afeta a comunicação dos dados diz respeito à garantia da privacidade dos pacientes que fornecem suas informações e amostras biológicas. No entanto, como explicado anteriormente, as pesquisas realizadas pelos entrevistados variaram significativamente em relação à natureza dos dados. A área de medicina, que trata diretamente com os dados dos pacientes é, portanto, diretamente atingida pelas questões éticas.

A pesquisadora P08 explicou que o seu projeto de pesquisa conta com uma central de dados. Todas as solicitações de acesso ao material passam por análise e, caso o pedido seja aprovado, os dados são encriptados de forma que a privacidade dos participantes seja garantida. Segundo definição da enciclopédia (INFOPÉDIA, 2016), a encriptação é um processo de criptografia dos dados, no qual algumas informações são convertidas para um código especial, de modo que sejam incompreensíveis para quem não tenha acesso a tais códigos. A finalidade desse processo na divulgação dos dados de pesquisa é permitir o acesso apenas à parte dos dados, excluindo-se assim aquelas informações que identificam o paciente.

P08: [...] Uma vez o projeto aprovado, o centro de dados manda o dado para a pessoa, encriptado. Tem as regras, né? Para evitar confusão e proteger o dado mesmo, proteger o participante, que não vem com o nome tudo bem, mas mesmo assim tem que ter cuidados e critérios.

O pesquisador P09, também da área de medicina, explica que os dados são referentes às amostras biológicas coletadas dos pacientes e que, para isso, é necessário identificá-las. Tal procedimento precisa ainda ser aprovado pelos comitês de ética, que exigem que cada participante assine um termo em que declara ter conhecimento de que seu material biológico fará parte de uma pesquisa.

P09: Veja bem, esse projeto tá no comitê de ética, porque cada pessoa que for participar tem que assinar uma declaração, um termo de TCE, um termo de consentimento esclarecido. Então eles têm que assinar isso daqui. Às vezes identifica, às vezes não é necessário se identificar. Por exemplo, eu estou orientando um trabalho agora que nós estamos fazendo um estudo dos óbitos por fatores associados aos óbitos por dengue no Distrito Federal. Nós fomos dispensados do termo de consentimento porque não vai aparecer o nome da pessoa, dos setenta e poucos óbitos que nós estamos estudando, vai aparecer só as características dele, não o nome, identificação nenhuma.

MC: Nessa outra pesquisa, sobre zika, é possível anonimizar os dados?

P09: Bom, em um projeto de pesquisa não, porque eu tenho que pegar dados de uma pessoa e ela precisa [assinar]. Não são dados secundários e aí ela precisa assinar, que permite que eu colha sangue, que eu armazene sangue.

Já na biologia e na farmácia, tais implicações não foram apresentadas como determinantes, visto que as pesquisas desenvolvidas nas duas áreas não lidam diretamente com os dados dos pacientes e, quando isso ocorre, os dados já estão em formato de anonimato. As duas áreas tendem a trabalhar com dados dos pacientes coletados em outras etapas, que já receberam aprovação do comitê de ética responsável e já passaram pelo processo de criptografia dos dados.

P10: Quando você tem alguma coisa feita com o paciente, você já tem um protocolo de ética possivelmente já aprovado pelas instituições competentes. Então isso não seria problema. O paciente em geral... Quando você disponibiliza alguma coisa relacionada a paciente, você não coloca nome nem indicação de quem seja.

Em todas as três áreas, é necessário que o projeto da pesquisa seja aprovado por comitês de ética. A exigência é colocada tanto por parte das agências financiadoras de pesquisa, como pelas políticas dos periódicos científicos. Portanto, os aspectos éticos da pesquisa precisam ser observados desde o início da pesquisa até a divulgação dos resultados. No entanto, tais aspectos podem ser impeditivos para que áreas como a medicina façam a comunicação aberta dos dados, mas não se configura como barreira para outras áreas como biologia e farmácia.

O impedimento para no campo da medicina relaciona-se diretamente com a discussão da categoria, que abrange a colaboração. Conforme explicou P08, o uso dos

dados coletados por outras pesquisas pressupõe um conhecimento amplo sobre o contexto dos dados. No entanto, para que os dados brutos das pesquisas sejam disponibilizados, são necessários processos de descaracterização dos dados que possivelmente identifiquem os participantes da pesquisa. Com isso, a ampla compreensão do contexto onde os dados foram coletados fica prejudicado, bem como seu reuso por outros projetos de pesquisa que não envolvam os pesquisadores que geraram os dados.

5.2.2.2. Subcategoria “condições de análise dos dados”

A disponibilização ou a possibilidade de acesso aos dados está diretamente relacionada com a capacidade de analisá-los. De forma geral, todos os pesquisadores das áreas de biologia e farmácia afirmaram que a análise dos dados suportada com os recursos ideais permite que os resultados sejam obtidos de forma mais rápida. No entanto, os pesquisadores não dispõem desses recursos ideais, principalmente quando comparado aos pesquisadores dos países desenvolvidos. Portanto, os pesquisadores entrevistados demonstraram receio em divulgar os dados de pesquisa à medida que esses vão sendo coletados, ou seja, antes de publicarem formalmente os resultados. Isso porque a disponibilização parcial dos dados poderia proporcionar que os resultados fossem alcançados por pesquisadores com melhores condições de acesso aos recursos necessários à análise. Os pesquisadores nessas condições foram recorrentemente descritos como oriundos dos países desenvolvidos, em contraste com os pesquisadores brasileiros. Com isso, os pesquisadores de instituições brasileiras estariam constantemente em desvantagem em relação às práticas de comunicação aberta e imediata dos dados.

O pesquisador P01 explicou que tais práticas de comunicação dos dados é vantajosa para a ciência, pois acelera os resultados das pesquisas. No entanto, ela é assimétrica no que diz respeito aos benefícios para os pesquisadores. Pesquisadores com amplo reconhecimento internacional e com melhores recursos para pesquisas conseguem mais colaboração em suas pesquisas, enquanto pesquisadores sem os mesmos acessos aos recursos correm o risco de ter sua “ideia roubada” e não serem reconhecidos por ela.

P01: Só que esses grupos são grupos muito estabelecidos, então provavelmente o que eles conseguem com isso é muita colaboração. Então, se eu sei que eles já estão fazendo isso, eu posso me juntar a esse grupo e a gente começa a fazer essas análises juntos. Então, no final

das contas, eles ganham muito mais do que perdem, mas se eu fizer isso, provavelmente, eu posso... não é uma coisa muito, como eu poderia dizer? Não é muito simétrica, entendeu. Principalmente país desenvolvido e o Brasil. Ainda é um pouco assimétrica essas coisas assim, sabe.

A desvantagem para os pesquisadores de países em desenvolvimento como o Brasil deve-se primordialmente às diferenças de recursos financeiros e de infraestrutura. Segundo P07, a superioridade em relação aos recursos financeiros coloca os países desenvolvidos com acesso aos melhores equipamentos, que lhes permite maior capacidade na geração e análise dos dados.

P07: Eles têm uma maior capacidade na geração desses dados, porque eles têm mais equipamentos, eles têm mais dinheiro.

A infraestrutura propiciada a partir dos recursos financeiros faz com que os pesquisadores brasileiros estabeleçam parcerias com laboratórios e grupos de pesquisa dos países desenvolvidos. Segundo P03, as parcerias têm por objetivo a utilização dos equipamentos dos grupos com melhor infraestrutura. Como um dos resultados da parceria, os dados da pesquisa são compartilhados com os pesquisadores possuidores da infraestrutura.

P04: Com outro grupo lá fora, porque tem isso já estabelecido no laboratório dele. Então, é mais fácil do que eu montar aqui. É mais barato eu ficar lá seis meses e trazer essas modificações pra depois voltar e trabalhar aqui. Porque montar aqui numa época que tá pobre o país, é complicado. Eu não teria dinheiro pra montar essa estrutura que eles têm lá já montado. Eles vêm que a gente faz junto.

MC: E se vocês, por exemplo, vão pra lá, igual a senhora falou, vão pra lá seis meses, coleta os dados, eles têm acesso a esses dados também?

P04: Qual, os meus? Como projeto, sim.

A infraestrutura ideal reúne os equipamentos e outros materiais necessários, tais como os reagentes. Nesse quesito, as pesquisas realizadas em países em desenvolvimento também se mostram em desvantagem, pois a qualidade dos reagentes que chega a tais regiões é inferior àqueles disponíveis em países desenvolvidos. Isso faz com que tenhamos resultados diferentes daqueles obtidos em melhores condições. Segundo P10 e P04, os reagentes são importados dos países desenvolvidos. Isso faz com que tenhamos que pagar valores mais altos para adquiri-los e esperar períodos de tempos maiores para recebê-los. O processo de transporte e a demora em utilizar tais reagentes pode ainda afetar sua funcionalidade, resultando na ineficiência de seu funcionamento.

P10: A gente tem o acesso ruim a reagentes e mesmo aos equipamentos. É ruim, demora muito. Os fornecedores brasileiros são péssimos. O mercado brasileiro é ruim. Então eles não têm estoque local ou cada vez que você precisa de um reagente é sessenta dias. Lá não dá uma semana, pensa na diferença. Então é muito diferente e às vezes chega aqui e você não sabe por que não funciona. Não funciona. Não funciona e você vai lá fazer a pesquisa, como muitas vezes eu fui, e disse: “poxa, faço a mesma coisa na França que eu faço no Brasil e lá dá certo e aqui não dá, por quê?” Porque o reagente que chega aqui, chega bichado. Desembarca lá não sei quando e vai e fica não sei quantos dias no aeroporto pra Anvisa liberar. Quer dizer, isso aí vai acabando com o sistema. Fora que para gente custa cinco vezes mais do que para eles. Além de eles terem mais dinheiro, para gente custa mais caro.

P04: A segunda coisa é que a gente sabe é que tem a demanda do material que a grande maioria do material é importado. Então, lá o material, na maioria dos centros de pesquisa, tem a manufatura desse material in loco. Então, o cientista desce no mesmo prédio, vai no subsolo ou, enfim. Já tem aquele material em mãos.

Nos termos apresentados pelos pesquisadores entrevistados, os países em desenvolvimento apresentam menos recursos para a análise dos dados quando comparados aos grupos de pesquisa dos países desenvolvidos. Tais condições fazem com que seja desvantajoso para os pesquisadores dos países em desenvolvimento comunicarem abertamente seus dados.

P01: Se eu disponibilizo essa informação e um grupo com um pouquinho mais de recurso, os caras conseguem fazer, em segundos hoje, numa semana, o que a gente aqui leva meses pra poder fazer. Então, a gente tem trinta amostras para poder fazer o genoma, a gente sofre de problema com importação. A fundação que compra as coisas leva mais ou menos um mês e meio para poder fazer uma compra. Então, se eu divulgo um resultado rapidamente que vai permitir que grupos, que são competidores (porque estão fazendo a mesma coisa), eles vão ter acesso muito mais rapidamente a essa informação e eles vão acabar publicando antes. Esse é um outro problema.

Apesar da deficiência em infraestrutura, por vezes os países em desenvolvimento são privilegiados no acesso ao material biológico, como no caso das pesquisas relacionadas ao vírus da zika. Com isso, os pesquisadores dos países desenvolvidos dependem de alguns dados oriundos de amostras biológicas obtidas aqui para a realização de suas pesquisas sobre o zika. Segundo o relato de P10, tem sido recorrente a visita de pesquisadores estrangeiros em centros de pesquisa sobre o vírus no Brasil. No entanto, o principal interesse desses pesquisadores é o acesso aos dados das amostras biológicas e não necessariamente uma colaboração científica de fato com os pesquisadores brasileiros.

P10: A gente tem recebido muita gente de fora, solicitando zika, chikungunya... O pessoal vem aqui e aí eles falam: "o que que vocês tão fazendo?". "A gente tá fazendo isso, fazendo aquilo". E eles: "aham, aham, aham". Até o momento que você fala assim: "olha, e aí a gente coletou umas amostras de paciente soro lá, não sei aonde". E aí eles falam: "Opa". Quer dizer, é isso que eles querem, né? Eles querem a amostra, eles querem o dado, eles querem essas coisas, porque o resto eles já têm.

5.2.2.3. Subcategoria "aspectos legais"

O impasse entre a ampla possibilidade de acesso aos dados brutos advindos de amostras biológicas dos pacientes e a falta de condições ideais para analisá-los alcançam a discussão no âmbito dos aspectos legais, que envolvem três fatores. O primeiro relaciona-se com o âmbito jurídico da legalidade em compartilhar alguns tipos de dados, tais como material genético. Conforme apresentado anteriormente, em 2015 entrou em vigor uma lei brasileira que regula o acesso ao patrimônio genético do país, a Lei nº 13.123/2015. A aplicação da lei resultou na impossibilidade de envio de amostras biológicas para o exterior, dificultando assim o desenvolvimento de algumas pesquisas fora do país, como, por exemplo, sobre o vírus da zika. Conforme relatado por P03, P04, P10 e na reportagem *Brasil Soneta* (2016), especialmente no caso das pesquisas sobre o vírus da zika, os pesquisadores estrangeiros têm demonstrado interesse em acessar os dados relacionados às amostras biológicas. Portanto, o âmbito jurídico se mostra como um dos fatores de influência nos aspectos legais da comunicação dos dados de pesquisa, especialmente a citada legislação brasileira.

O segundo aspecto também está emerso no campo jurídico, no entanto trata das garantias acerca da propriedade intelectual. Segundo a declaração de P10, o desenvolvimento tecnológico, que no caso do vírus da zika é o soro e a vacina, demanda proteção intelectual como condição para que a tecnologia seja transformada em produto. Nesse caso, a proteção permitiria expectativa de lucro com a exploração do conhecimento produzido. Portanto, a possibilidade de a pesquisa resultar em produto lucrativo, distanciaria os pesquisadores das práticas da ciência aberta

P10: O nosso enfoque é o desenvolvimento de tecnologia, é o desenvolvimento de soro e vacina contra o zika. Já nessa área a coisa não é tão *open science* assim. Porque, quando você faz pesquisa de tecnologia, você tem que pensar em proteção intelectual para aquilo virar um produto. Porque sem a proteção intelectual o produto perde a lucratividade. O interessante então nesse sentido... Se você fizer mesmo uma busca, você vai ver que, na questão tecnológica de zika, o pessoal segura mais.

Observa-se ainda que a forma como a exploração comercial é empenhada baseia-se legalmente nos direitos de propriedade intelectual. Segundo Diniz (2005), tais

direitos são regulações jurídicas de um produto imaterial, a informação e a expressão. Historicamente estão relacionados com a emergência da produção industrial e com o desenvolvimento de um sistema econômico baseado no lucro (BRANCO, 2007).

Nesse contexto, são utilizadas as patentes como forma de limitar o acesso ao conhecimento científico e tecnológico com vistas aos interesses comerciais e industriais. Isso porque as patentes são instrumentos para garantir exclusividade de produção e comercialização de um invento (CASTRO, 2012). Entretanto, nem a patente nem os direitos de propriedade intelectual de forma mais ampla são direitos absolutos, conforme defendeu Pinto (2009). A propriedade intelectual, bem como as propriedades de qualquer natureza devem exercer uma função social, além de beneficiar seu proprietário. No caso da produção industrial de fármacos, o acordo TRIPS, firmado em 1994, é um instrumento legal internacional que garante que países pobres obtenham uma licença do direito de propriedade intelectual e, assim, possam produzir versões genéricas de fármacos destinados a graves acometimentos de sua população (WORLD, 1994).

5.2.3. Categoria “financiamento”

A categoria “financiamento” abarca duas perspectivas: a infraestrutura necessária aos processos de comunicação dos dados e os interesses dos grandes financiadores de pesquisas nesse campo. A infraestrutura foi destacada por P07 como condicionante para a abertura dos dados na comunicação científica. Como forma de sustentar seu argumento, o pesquisador apresentou o projeto inglês de pesquisa sobre o vírus da zika, que desenvolveu um portal onde são compartilhados os dados coletados em tempo real, ou seja, à medida que vão sendo processados.

P07: Uma iniciativa lá da Inglaterra na verdade, onde eles montaram esse site. Eles vieram aqui no Brasil, ficaram dois meses coletando amostras e fazendo o sequenciamento. Eles tinham um projeto aí de sequenciar mil genomas diferentes aqui do Brasil. E o legal, todos os protocolos, tudo que eles fizeram eles disponibilizaram em tempo real.

O pesquisador considerou que tal iniciativa só foi possível em razão da infraestrutura que o projeto dispõe. Em relação à experiência do projeto inglês, P07 considera necessário para a divulgação dos dados da sua pesquisa, com a mesma rapidez e abertura, a existência de um banco de dados adequado, relacionado com o

financiamento da pesquisa. Para o pesquisador, a estrutura do sistema de informação seria um instrumento para a agência de fomento gerenciar o desenvolvimento do projeto de pesquisa, sua qualidade e sua divulgação.

P07: A gente tinha que ter sobretudo uma estrutura pra poder analisar essa divulgação dos dados em tempo real. Como é que esse pesquisador, ele vai fazer essa divulgação? Ele vai ter um portal próprio, [...] um banco de dados, por exemplo, vai suportar o que vai ser desenvolvido pela própria agência de fomento. Porque daí ficaria mais fácil também ela regulamentar e acompanhar essa divulgação. Tu precisarias também de um próprio técnico pra ver se aqueles dados que estão disponibilizados estão corretos ou não. O cara tá disponibilizando ali dados, porque eu posso, por exemplo, disponibilizar dados de materiais e métodos da forma que eu vou fazer ou dizer que eu coletei animais. Lógico que eu coletei número X de animais, são dados que vão encher linguiça, mas não são resultados propriamente ditos. Porque ainda eu não sequenciei, porque sequenciar ainda, principalmente sequenciar o genoma inteiro, é algo caro. É uma tipologia que ainda está na mão de poucas pessoas. Exige então um conhecimento técnico específico, exige recursos. Então muitas pessoas podem coletar as amostras e dizerem que vão fazer e não fazer de fato, então também tem essa questão, que isso tem que ser regulamentado. “Ah, fulano de tal ele coletou sei lá, mil amostras”. Tá, mas ele não sequenciou nenhuma. Não é possível que ele coletou cem amostras e já sequenciou dez. Essas questões também têm que ser analisadas.

Entre os elementos elencados por P07 também se destaca a necessidade de um profissional que possa fazer as mediações necessárias para o sistema de informação. Uma de suas atribuições seria avaliar a qualidade dos dados que estão sendo apresentados como resultado da pesquisa financiada.

O outro elemento diretamente relacionado ao financiamento das pesquisas diz respeito aos interesses envolvidos. Em especial, foi destacado o interesse dos Estados Unidos em relação às pesquisas sobre o vírus da zika. O interesse do país tem resultado em altos financiamentos para pesquisas sobre o tema. Para alguns pesquisadores, como P09 e P10, essa atenção recebida faz com que as doenças causadas pelo vírus da zika não se enquadre como negligenciada, apesar de serem transmitidas pelo mesmo vetor causador de doenças classificadas como negligenciadas. O principal elemento que justifica a classificação relaciona-se com os investimentos aplicados em pesquisa, especialmente de financiadoras norte-americanas e de grupos da indústria farmacêutica.

P09: Eu acho que ainda não está negligenciada porque ainda está com um nível de impacto muito grande e o investimento muito alto. O que que é negligenciar? São aquelas doenças cujo controle não estão recebendo os recursos financeiros suficientes. Pra você ter uma ideia, eu faço parte do comitê de avaliação de um método de aprimoramento do controle vetorial do *Aedes* de uma bactéria chamada Wolbachia, uma técnica australiana, tá sobrando dinheiro, o ministério [da saúde] deu, a Fap do Rio de Janeiro deu, porque nós estamos fazendo no Rio de Janeiro, Tubiacanga, Jurujuba e agora vai ser em Niterói, depois Belo Horizonte. A Fundação Bill Gates

tá financiando, olha, então não pode ser negligenciada, tá sobrando dinheiro. Outro dia eu tava falando com ele, vamos ver quem vai financiar. “Ah não precisa, nós já temos o financiamento.” O que é doença negligenciada? É doenças do pobre? Não. Doenças negligenciadas são aquelas doenças que não estão recebendo recursos suficientes. Agora, ela não pode ser negligenciada, porque os Estados Unidos estão arrepiados. O Obama pediu um bilhão e oitocentos mil dólares no Congresso americano para pesquisa nessa área de zika, vacina e controle vetorial e tratamento.

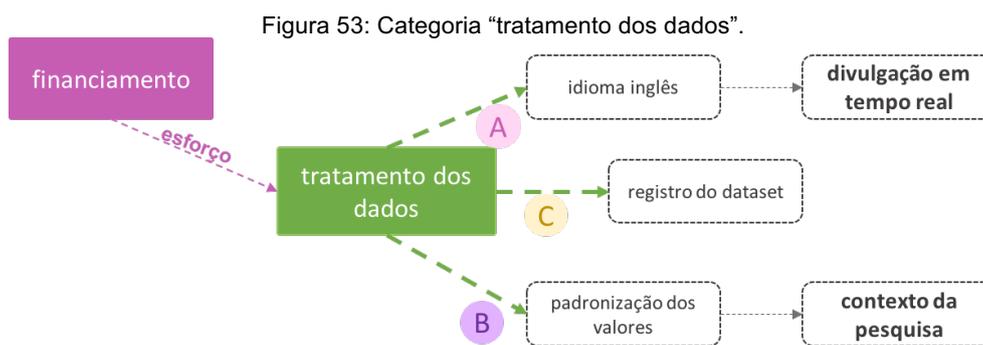
Segundo P09, uma das grandes fontes de financiamento é de origem direta do governo dos Estados Unidos da América. O país teria iniciado o investimento devido aos casos de pessoas contaminadas pelo vírus, sobretudo na Flórida. Com a possibilidade de expansão do vírus para o país, o governo teria destinado mais de um bilhão de dólares para pesquisas. A possibilidade de expansão do vírus para países ricos teria ainda despertado o interesse da indústria farmacêutica para o desenvolvimento de fármacos, conforme explicou P10. Para o pesquisador, a expectativa de lucro com a comercialização de medicamentos foi decorrente da contaminação de pessoas de países desenvolvidos e da possibilidade de expansão do vírus para esses países. No entanto, caso a dispersão do vírus não se confirmasse o interesse da indústria farmacêutica poderia desaparecer.

MC: A Zika está despertando mais interesse da indústria farmacêutica para produção do medicamento?

P10: Eu acho que no começo sim. A gente não sabe o que vai acabar acontecendo, porque, dependendo da evolução do espalhamento e da dispersão do zika, eles podem perder o interesse. Não é novidade, lógico que eles trabalham atrás de lucro e só vai investir em coisa que seja lucrativa. Na história do zika, tem um pouco de mobilização inicial, porque tem essa, esse componente da transmissão sexual que eu te disse que o transforma em um vírus que pode ter uma circulação mais ampla dentro do mundo, atinge todos. Porque a transmissão sexual atinge todos os países.

5.2.4. Categoria “tratamento dos dados”

O tratamento necessário para a comunicação aberta dos dados foi um fator apresentado como inibidor para o processo, uma vez que o tratamento demandado é mais dispendioso do que aquele necessário para os fins apenas da pesquisa. As dificuldades variam em função da área do conhecimento, sendo assim possível observá-las em distintos níveis (Figura 53).



Fonte: Elaboração própria.

Segundo P10, na área de farmácia, os dados coletados no laboratório são registrados em um caderno de anotação, já que é proibida a entrada de computadores nos laboratórios onde é gerada uma grande quantidade de dados. Apenas quando finalizam o registro dos dados no caderno é que alguns dados são selecionados para serem transcritos para o computador. A seleção se dá de acordo com a análise desejada para um projeto de pesquisa específico e suas expectativas de publicação dos resultados. O pesquisador P10 explicou que, caso os dados brutos fossem amplamente comunicados, todos os registros do caderno deveriam ser transcritos para um arquivo digital, o que demandaria mais esforço e tempo para a atividade.

No campo da biologia, os dados também precisariam ser registrados em sua totalidade, caso haja a expectativa de seu reuso, conforme relatou P01.

P01: Na verdade, eu acho que o compartilhamento dos dados demanda que a pessoa que recebe esse dado no final, ela consiga fazer de maneira autônoma. Então, se eu vou colocar uma planilha, eu vou investigar ao máximo se uma outra pessoa vai conseguir... Porque, às vezes, eu tô fazendo uma planilha, quando é pra mim, eu vou lidar com ela, muitas vezes, como se ela tivesse nesse estado mais bruto. Eu vou coletar as informações que eu preciso, mas se uma outra pessoa vai... eu vou ter que organizar melhor esse dado, para que a pessoa, autonomamente, consiga acessar a informação e que isso seja útil. Acho que esse é o ponto. Porque, às vezes, não adianta você só disponibilizar, mas a outra pessoa não conseguir entender como foi feito, porque você tem que disponibilizar tudo: como foi feita a coleta; como você organizou os resultados; que tipo de processamento você fez. Que são coisas que, às vezes, a gente vai descrever só no final da pesquisa, quando você já tá escrevendo o artigo, entendeu? Se você for compartilhar, você já precisa fazer isso imediatamente, enquanto você tá coletando os dados. Você tem que ter essa regra, porque a hora que você acabou de coletar, isso já vai ter que tá disponível. Então, acho que isso ajuda a organizar um pouco melhor o trabalho.

A necessidade de registro e descrição da totalidade dos dados é apresentada por P01 como um requisito para a disponibilização aberta dos dados. No entanto, diferentemente do pesquisador da área de farmácia, espera-se que esse processo seja realizado à medida que os dados sejam coletados, para que estejam imediatamente

disponíveis. Observa-se ainda que isso seria apenas um adiantamento de um processo que já é necessário para a publicação dos resultados. Portanto, entendeu-se que esse não é o principal problema relacionado ao tema. Segundo P01, a dificuldade relaciona-se com a necessidade de realizar todos os registros e descrições em inglês.

P01: Primeira coisa para mim, é que eu teria que já... isso não é um empecilho, mas eu já teria que organizar isso tudo em um idioma: o inglês. Então, todos os meus protocolos, todos os meus dados já teriam que estar organizados em inglês. Se eu não fiz desde o início, então eu vou ter que ter esse... se eu quiser compartilhar, eu vou ter que ter essa etapa para poder colocar tudo em inglês. Então, isso seria um trabalho a mais, que eu teria que fazer, para poder compartilhar. Ou você já decide que você vai organizar toda a informação em inglês desde o início. Esse é um empecilho.

Já para o campo da medicina, o tratamento dos dados abarca as problemáticas relacionadas à padronização dos dados. Segundo P08, os valores dos dados coletados variam em função da região ou grupo que está sendo estudado. Além disso, em alguns casos as variáveis previstas para um contexto de pesquisa não são aplicáveis em outras experiências. A pesquisadora exemplificou com o relato de uma pesquisa realizada primeiramente com a população do Nordeste brasileiro e depois no Sudeste. As possibilidades de respostas previstas para a região nordestina não contemplavam as práticas no sudeste. Com isso, quase metade dos respondentes selecionaram uma resposta genérica e que não representavam suas práticas, denominada “outros”. Após investigações, os pesquisadores descobriram o que o valor “outros” de fato representava para a região e os dados tiveram de ser corrigidos. Portanto, P08 afirma que a disponibilização desses dados brutos precisa estar estruturada de forma concisa e deve ser acompanhada de um dicionário de variáveis que seja capaz de possibilitar o entendimento dos dados e evitar ambiguidades.

P08: Eu tenho que fazer um dicionário bom de variáveis, um dicionário de variáveis decente. Eu tenho que recategorizar variáveis, porque a gente quando faz a pergunta, a gente se baseou muito nas grandes pesquisas nacionais, IBGE, FINAD não sei o quê. Para ter um questionário validado. Tem coisa que não funciona no Rio de Janeiro e tem coisas que a gente descobre a caminho e que a gente codifica a posteriori. Então tem todo um tratamento que você tem que dar. Quer um exemplo?

MC: Uhum

P08: Dada a água que você bebe, o que você faz com a água de beber? Filtra, ferve, coa? Ninguém coa no Rio de Janeiro, mas no nordeste se coa a água. Porque vem com barro, dependendo do lugar. Então, filtra, ferve, coa, não faz nada e outros. Você sabe o que foram o “outros”?

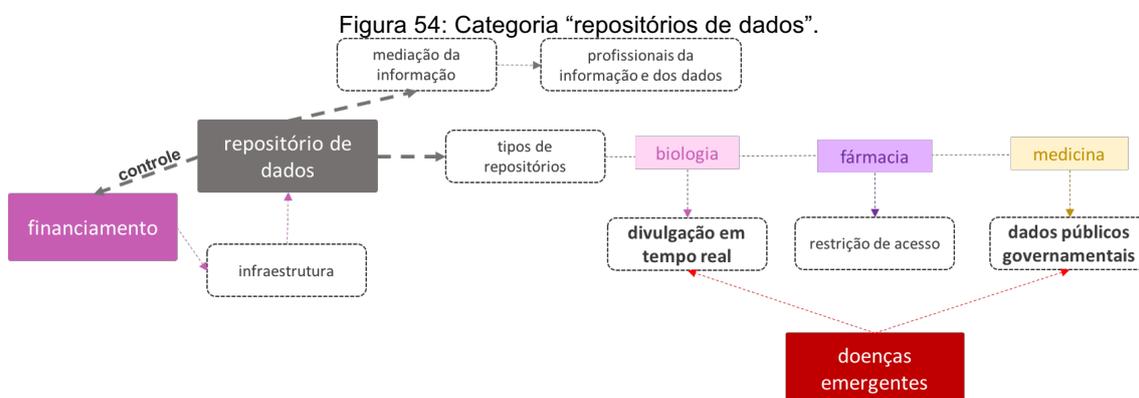
MC: O que?

P08: Galão de água. Compra garrafão, aquele galãozão, que na verdade sei lá, 40% das famílias estavam comprando água em Manguinhos, por incrível que pareça. Porque a água estava

fedorenta, sabe? Então a gente recriou uma categoria, mas isso foi a posteriori. Você olha o “outros” vê lá “Galão, galão, água mineral, galão”, aí você tem que arrumar isso e criar uma categoria pras pessoas analisarem. Claro que dá trabalho. Limpar banco de dados pra viabilizar um “outro” analisar é um trabalho grande, não é pequeno não.

5.2.5. Categoria “repositórios de dados”

Os elementos pontuados pelos pesquisadores a respeito dos repositórios de dados também podem ser sistematizados em função das três áreas do conhecimento aqui discutidas (Figura 54). Inicialmente, é relevante destacar que nenhum dos pesquisadores entrevistados usou o termo “repositórios de dados”, e sim banco ou base de dados. No entanto, conforme a discussão apresentada na revisão de literatura dessa tese, o termo repositório de dados engloba a noção de banco e base de dados, estando ainda de acordo com os propósitos da ciência aberta.



Fonte: Elaboração própria.

Segundo Walport e Brest (2011), os repositórios de dados são uma estratégia eficiente para a realização da comunicação dos dados de pesquisa no contexto da ciência aberta. Eles são destacados pelos principais projetos europeus de promoção da ciência aberta como infraestrutura necessária para a gestão e disponibilização dos dados de pesquisa, conforme indicaram Pampel et al. (2013). Portanto, além da infraestrutura, representada por todos os elementos necessários ao desenvolvimento do repositório, destaca-se também o aspecto da gestão, envolvida diretamente com a manutenção dos sistemas de informação. A gestão é realizada pelos pesquisadores, por meio do tratamento dos dados, e por profissionais, por meio de serviços de informação e apoio aos pesquisadores.

O destaque da estrutura como necessária à comunicação dos dados de pesquisas e às práticas de abertura e colaboração da ciência foi feito especialmente pelos pesquisadores relacionados à biologia. Tais pesquisadores já utilizam repositórios de dados internacionais, especialmente para o compartilhamento dos sequenciamentos dos genomas estudados. Em decorrência de suas necessidades, os pesquisadores consideraram que a existência de sistemas dessa natureza é importante para auxiliar em suas pesquisas, bem como no gerenciamento e na divulgação dos dados que produzem, tal como previsto por Walport e Brest (2011).

P01: A gente precisa para fazer isso... ou eu coloco em um banco de dados que já existe... Eu realmente queria que a universidade ajudasse, tivesse uma infraestrutura, onde a gente tivesse um site, que a universidade criasse, a gente pudesse manter. Seria muito legal eu fazer isso. Se eu quiser, ou eu mesmo faço ou pago uma empresa. Só que a gente não tem dinheiro para fazer isso, então isso vai para um banco de dados público qualquer, que não vai trazer visibilidade imediata para a universidade, entendeu? Porque se você tem um site onde as pessoas que querem utilizar essa informação elas entram, elas vão conhecer a universidade, elas vão ver que tipo de pesquisa que tá sendo feita aqui. Se eu envio para um outro banco, só é mais uma sequência naquele lugar, entendeu. Então, essa é uma coisa que eu gostaria muito.

O banco de dados citado por P01 é o repositório National Center for Biotechnology Information (NCBI)⁵⁰, que é mantido pelo governo estadunidense. O sistema teve início em 1998, como um produto do National Library of Medicine (NLM), vinculado ao NIH. O objetivo inicial do NCBI era reunir, tratar e disseminar os dados genômicos. Posteriormente, passou a abranger para vários temas da biologia molecular (NATIONAL, 2017). O repositório, além de receber depósitos dos dados, ainda apresenta ligação com outros repositórios da mesma natureza, portanto pode ser classificado como agregador de conteúdo ou *meta database* (ARAKI et al., 2012).

A fonte provedora de dados para o agregador são os repositórios de universidades, institutos de pesquisas e centros de dados de vários países do mundo. Portanto, P01 considera necessária a existência de um repositório em sua instituição, de modo a facilitar o seu trabalho e ampliar a visibilidade da própria instituição. Tal imperativo também foi apontada por P02, que denominou o sistema de central de dados. Embora a nomenclatura aponte para um sistema amplo, P02 restringiu sua funcionalidade aos pesquisadores vinculados ao laboratório mantenedor da base, reafirmando assim a proposta de P01.

⁵⁰ Fonte: <https://www.ncbi.nlm.nih.gov/>

P02: [00:14:45] Eu acho que esses dados primários não teriam nenhum problema, não teria nenhum problema. Desde que... então, o que que acontece no meio da pesquisa? É porque a gente disponibiliza os dados, mas, assim, a gente teria que ter uma central, né? Que a gente... na aids, por exemplo, aqui tem uma central de dados, eles chamam de “central de dados”. Então você centraliza todos os dados lá e todos os pesquisadores que trabalham pra aquele laboratório, ele tem acesso a todos aqueles dados, pra não ficar coletando novamente.

A existência do repositório representa ainda uma forma de compartilhar a responsabilidade da gestão dos dados com a instituição que deseja acessá-los, seja a financiadora ou o próprio pesquisador. Considerando que a comunicação dos dados de pesquisa acrescenta mais esforços no tratamento dos dados e que os repositórios são um instrumento para gerenciá-los, P02 enfatiza que as instituições também precisam assumir parte do papel na gestão dos dados, oferecendo assim infraestrutura e serviço de informação especializado.

P02: Eu acho que para gerar isso tem que se criar, é aquilo que eu falei, uma central de dados, onde tivesse gerentes dessa central de dados, para fazer tudo direitinho. Porque, assim, no Brasil, por exemplo, no nosso laboratório: eu faço tudo, a gente faz tudo então, eu sou a pesquisadora, chefe do laboratório; eu tenho que escrever o projeto; eu tenho que veicular os dados; eu tenho que ver se os dados estão certos; eu tenho que ir para bancada ver se tá bom; eu tenho que fazer tudo. Então, é impossível você criar um banco direitinho e disponibilizar isso.

A expectativa dos pesquisadores com a oferta dessa infraestrutura e serviço em suas instituições é facilitar o processo de registro e disponibilização dos dados, conforme explicou P01.

P01: Uma coisa é a dificuldade, às vezes, da gente conseguir isso na própria universidade. Tipo assim, eu não preciso disponibilizar meu banco de dados, eu poderia disponibilizar em alguma coisa aqui dentro da universidade mesmo, que a universidade ela tivesse uma infraestrutura (às vezes até tenha, a gente desconheça), para você poder fazer esse tipo de coisa de uma maneira mais simples, sabe, aonde você pudesse alimentar isso.

De forma geral, os repositórios citados por P01 e P02 disponibilizam amplamente os dados registrados, segundo as expectativas da ciência aberta. Na proposta levantada por eles, não se notou nenhum desvio nesse padrão de abertura. Segundo a experiência dos pesquisadores, os repositórios internacionais não apresentam restrição de acesso aos dados que são obrigatórios para a publicação dos resultados das pesquisas em artigos de periódicos.

P07: A gente tem um banco de dados chamado de Ibank, que é um banco de dados emitido pelo governo americano. É uma estrutura onde estão todas os genomas de todas as espécies de micro-organismos, espécies maiores são depositadas lá. É um banco de acesso livre, então qualquer pessoa pode acessar lá e pegar esses dados. As revistas, antes de você publicar um paper, te obrigam, entre aspas, que você faça o depósito dessa sequência para que de fato ela fique disponibilizada.

Diferentemente dos pesquisadores no campo da biologia, os pesquisadores da área de farmácia pontuaram restrições necessárias para os repositórios de dados. É importante ressaltar que os pesquisadores vinculados às áreas de farmácia e medicina não possuem prática de registro e disponibilização de seus dados em repositórios, de nenhuma natureza. Entretanto, no processo da entrevista foi solicitado que eles descrevessem os requisitos necessários para um repositório em seu campo de pesquisa, de modo que os motivassem a registrar e compartilhar seus dados. Nas duas propostas apresentadas pelos pesquisadores no campo da farmácia, observaram-se os mesmos requisitos: limite nacional e restrição de acesso. Vale ressaltar que as propostas foram rascunhadas por pesquisadores de diferentes grupos de pesquisa, em momentos diferentes, em que um não teve acesso à resposta do outro.

Os pesquisadores P04 e P10 enfatizaram que, caso registrassem seus dados em um repositório, o sistema deveria servir para a colaboração entre os pesquisadores brasileiros e, para tanto, deveriam contar com ferramentas de controle de acesso. P04 ainda acrescentou que os pesquisadores participantes deveriam assinar um termo nos quais se comprometessem a compartilhar todos os seus dados e proteger os dados dos colegas que tenham dados disponíveis no sistema.

P04: Mas se tivesse, realmente, se a intensão for montar um grande banco de colaboração nacional e que seja semiaberto. É semiaberto, ou seja, não aberta ao público em geral ou pra... mas que seja, que tenha de certa forma uma proteção ou, por exemplo, que os participantes tenham uma senha, que possam acessar, que... alguma coisa que, de certa forma... e que todos os envolvidos assinem um termo se comprometendo não só a declarar que todos abrissem os dados, como também proteger os dados do colega. Porque aí também, imagina, né?

Infere-se do contexto apresentado que as preocupações relacionadas às restrições de acesso aos dados relacionam-se com a competitividade presente nas pesquisas do campo da farmácia. Tal aspecto não tem favorecido práticas de colaboração entre os pesquisadores, e suas limitações expressam-se também no desenho do sistema ideal.

Nas pesquisas relacionadas ao campo da medicina, observou-se que existe o amplo uso de bases de dados produzidas pelo governo sobre a população e a incidência de doenças. Conforme explicou P09, o Ministério da Saúde mantém sistemas de informação específicos para várias doenças e em diversos contextos. Tais sistemas apresentam informações abrangentes sobre o tópico abordado e são utilizados na análise dos resultados das pesquisas. Portanto, nessa área observou-se o uso de dados coletados por outros atores, mas que não são resultantes de pesquisa.

P09: Eu acho que quem produz o dado quer usar o dado e a originalidade, uma das características do artigo também. Eu nunca vi isso na minha área, alguém falar: “olha, eu tenho um banco de dados aqui”. A não ser os bancos de dados oficiais, dos sistemas de informações do ministério que são divulgados.

5.2.6. Categoria “reconhecimento e recompensa”

A principal motivação para o desenvolvimento das pesquisas, bem como para a divulgação dos seus resultados, é a expectativa dos pesquisadores em serem reconhecidos e recompensados pelo seu trabalho. Todos os entrevistados apontaram que as publicações acadêmicas são a expressão mais essencial do reconhecimento. Elas demonstram prestígio dentro do meio acadêmico e, apesar de serem independentes das diferenças disciplinares, são fruto da cultura acadêmica advinda da ciência moderna. Isso porque, conforme previu Garvey (1979), para que uma nova ideia seja considerada relevante na ciência, ela precisa ser divulgada e aceita por outros pesquisadores como um conhecimento científico válido. Com isso, o processo de divulgação torna-se essencial para a obtenção do reconhecimento almejado pelos pesquisadores. O principal meio de divulgação dos resultados apontados pelos pesquisadores das três áreas são os periódicos acadêmicos.

O periódico acadêmico representa uma importante tipologia das publicações científicas, que vêm evoluindo desde o século XVII, conforme o relato de Meadows (1999). No contexto analisado, os periódicos representam o principal meio de obtenção de status acadêmico entre seus pares, instituições e agências financiadoras de pesquisa. Portanto, ao mesmo tempo que geram reconhecimento e recompensa pelo trabalho realizado, as publicações acadêmicas são valorativas como instrumentos para obtenção de novos financiamentos.

A relação de dependência entre o reconhecimento e os periódicos acadêmicos foi determinada nas três áreas do conhecimento: A, B e C. Entretanto, a relação dos periódicos com a comunicação aberta dos dados se deu em função das diferenças disciplinares geralmente declaradas nas políticas editoriais dos periódicos. Nos casos em que a políticas editoriais dos periódicos requerem a comunicação aberta dos dados de pesquisa, pressupõe-se a necessidade de repositórios de dados de pesquisa.

A expectativa em divulgar os resultados da pesquisa em periódicos acadêmicos aponta para outro fator interveniente no processo, tratado aqui como a subcategoria “sigilo da pesquisa”. O sigilo, portanto, é visto como um pressuposto para a publicação em periódicos acadêmicos, que demandam originalidade e ineditismo. No campo da farmácia, o sigilo ainda foi apresentado como condição para a produção de tecnologias que serão patenteadas e tornar-se-ão objeto de exploração exclusiva em um determinado período temporal. Nas três áreas, o sigilo da pesquisa pode comprometer o fluxo da informação disponível e a velocidade com que ela é apresentada. O comprometimento com a disponibilização dos resultados de pesquisa e dos seus dados vai de encontro ao estabelecimento das ações coordenadas no contexto de doenças de emergência pública.

O interesse dos pesquisadores em serem reconhecidos e recompensados pelos esforços de sua pesquisa apresentou-se como elemento central entre os fatores que influenciam os dados de pesquisa. Essa foi a única categoria que se relacionou diretamente com as demais categorias. A relação foi estabelecida por meio das ligações com as subcategorias e propriedades da categoria “reconhecimento e recompensa” (Figura 55).



A categoria reconhecimento e recompensas derivou duas subcategorias que evidencia outros fatores relevantes para suas discussão. Suas relações são descritas e analisadas a seguir.

5.2.6.1. Subcategoria “periódicos acadêmicos”

Todos os pesquisadores entrevistados apontaram para os artigos de periódicos acadêmicos como o produto da comunicação de seus resultados que gera reconhecimento acadêmico. As publicações geram reconhecimento primeiramente pelos pares, que, ao mesmo tempo que os certificam como autores daquele resultado, promovem status acadêmico. Uma vez obtido o reconhecimento pelos pares, os autores das publicações poderão ter seus esforços de pesquisa reconhecidos pelas suas instituições e agências de fomento. É o conjunto desses reconhecimentos que representam as recompensas almejadas pelos pesquisadores. Portanto, os pesquisadores esperam ser recompensados pelos seus esforços, de forma que o reconhecimento gere status acadêmico, favorecendo o financiamento de seus projetos de pesquisa e a promoção na carreira de pesquisador.

Em todas as entrevistas, a expectativa de reconhecimento pelos dados produzidos foi ponto comum. No entanto, a singularidade das narrativas ocorreu, novamente, em função das diferenças disciplinares. Observou-se que a cultura acadêmica em torno da comunicação dos dados de pesquisa em artigos de periódicos acontece de duas formas distintas. Uma delas pode ser classificada como mais aberta e ocorrem na área das ciências biológicas, especialmente nos estudos que produzem dados sobre genoma. Nesse caso, as políticas editoriais dos periódicos tendem a exigir que os genomas sequenciados, utilizados como base para a proposição dos resultados, sejam integralmente disponibilizados em bases de dados. Tais bases são especializadas na tipologia específica dos dados e permitem que os artigos de periódicos as referenciem de forma adequada.

Os pesquisadores P01 e P07 explicaram que os genomas sequenciados normalmente são publicados em bases de dados, aqui denominados de repositórios de dados, quando um artigo é submetido para a publicação. Portanto, os dados sequenciados não são publicados à medida que são gerados, mas apenas com a expectativa de publicação do artigo.

P01: Antes de publicar não. Normalmente o que acontece é assim: na minha área (eu trabalho com genoma humano) qualquer trabalho que eu publique, eu preciso colocar todas as minhas sequências disponíveis em um banco de dados públicos. Então, eu não consigo publicar em nenhuma revista se eu já não tiver submetido essa informação. Algumas vezes eu tenho que colocar o dado final e o dado bruto. Algumas vezes. Outras não, só preciso colocar o dado analisado. Mas agora algumas revistas estão exigindo que você disponibilize o dado bruto em qualquer trabalho, uma planilha, o que for, você tem que disponibilizar, que é para que outros pesquisadores, eles consigam refazer e reanalisar e reutilizar sua pesquisa.

P07: A gente tem um banco de dados chamado de Ibank, que é um banco de dados emitido pelo governo americano. É uma estrutura onde estão todas os genomas de todas as espécies de micro-organismos, espécies maiores são depositadas lá. É um banco de acesso livre, então qualquer pessoa pode acessar lá e pegar esses dados. As revistas antes de você publicar um paper com fragmentos do genoma, elas te obrigam, entre aspas, que você faça o depósito dessa sequência para que de fato ela fique disponibilizada.

Na fala de P01, é possível perceber que o comportamento da disponibilização do conjunto dos dados, embora recorrente no campo da biologia, é recente. A tendência também foi apontada por P08, vinculada à área da medicina, mas com menor expressividade. A pesquisadora P08 também é editora de dois importantes periódicos na área da saúde pública, conforme descrito na sua apresentação (seção 5.1.7). A pesquisadora explicou que existem algumas linhas editoriais que estão exigindo a disponibilização do conjunto dos dados resultantes da pesquisa junto com o artigo. No entanto, a prática não parece tão sedimentada em comparação ao campo da biologia.

P08: Tem várias revistas que estão exigindo, toda a linha da PLOS tem ajudado. Eu sou editora associada na PLOS Neglected Diseases, e ela não é a que mais exige não, mas a PLOS One exige. A gente publicou um estudo do [A] lá e foi uma ginástica, porque teve mais juntar tudo [todos os dados].

Na experiência relatada por P08, identificou-se que existem algumas demandas para a disponibilização do conjunto de dados, no entanto a prática não é dominante. Em nenhum dos dois periódicos em que a pesquisadora compõe o conselho editorial, exigiu-se de maneira sistemática a disponibilização do conjunto de dados para publicação dos artigos, nem mesmo naquele participante da linha editorial citada no exemplo. A incipiência da discussão no campo da medicina pode ser comprovada pelas afirmações feitas por P09. Segundo o seu relato, os dados resultantes de uma pesquisa são divulgados apenas nos artigos, depois de selecionados e analisados. O pesquisador ainda afirmou que desconhece qualquer outra forma de comunicação dos dados além das publicações tradicionais.

P09: Não, não tem outro espaço para isso. Não existe, quer dizer, as revistas só aceitam é o artigo, sob forma de artigo.

A cultura da divulgação dos dados no campo da medicina assemelhou-se à da farmácia. Segundo o relato dos pesquisadores P03, P04 e P10, os dados são divulgados exclusivamente por meio de publicações científicas tradicionais, como os artigos de periódicos e patentes.

MC: Você já teve essa experiência de comunicar os dados?

P10: Não. [Só] o artigo mesmo.

A pesquisadora P03 ainda acrescentou que existe a expectativa de participar das publicações de artigos de periódicos como forma de ser recompensada pela colaboração na coleta e análise dos dados.

P03: Eu já vi isso, eu já passei resultado citotoxicidade, depois as pessoas publicaram, botaram o resultado que eu fiz. Eu gastei tempo, gastei material. Agora quando eu faço esse tipo de trabalho pra alguém, porque às vezes eles são enrolados com o tempo em termo de célula, eles não conseguem manipular direito a célula, mata a célula, contamina a célula, perde a célula, eu digo: “eu faço, mas com uma condição: o meu nome tem que tá no paper”.

5.2.6.2. Subcategoria “sigilo da pesquisa”

A publicação dos resultados da pesquisa em um artigo de periódico foi apontada pelos pesquisadores como a principal forma de obter reconhecimento pelos esforços empenhados na pesquisa. Uma das condições para essa publicação é que o seu conteúdo seja original e inédito. Portanto, o ineditismo foi apresentado como uma condição para obtenção de reconhecimento e recompensas no âmbito acadêmico. Conforme explicou P02, não é apenas o interesse em comunicar os resultados da pesquisa que está em jogo, e sim fazê-lo antes que qualquer outro pesquisador. Com isso, manifesta-se o receio dos pesquisadores em disponibilizarem os seus dados de pesquisa e perderem o ineditismo dos resultados de sua pesquisa.

P02: [...] se você disponibiliza seus dados antes de você publicar, isso cria... não torna mais inédito o que você tá encontrando. Isso é um problema, né? [...] Os problemas de você divulgar? É isso: porque você perde o ineditismo do que você está procurando. Infelizmente a ciência ainda é assim, né?

A necessidade de comunicar primeiramente os resultados da pesquisa foi apontada como empecilho por todos os pesquisadores entrevistados. Mesmo que o conjunto integral dos dados seja disponibilizado apenas após a publicação do artigo, alguns pesquisadores apontaram prejuízo em seus trabalhos. Como alternativa à problemática, P08 propõe um período de exploração exclusiva dos dados por parte dos pesquisadores que os produziram. Com isso, esse grupo poderia obter seu reconhecimento pelo esforço empenhado na pesquisa.

P08: No compartilhamento dos dados, eu acho que entra o: "Eu quero publicar primeiro". Se eu cedo o meu dado, assim que a pesquisa acabou, então normalmente quem defende, defende um período de propriedade, entendeu? É, o dado livre depois de dois anos. Porque eles precisam de recompensa pros dez anos que eles investiram e organizaram aquele raio daquela pesquisa. Eu entendo. É muito investimento, é chato pra caramba. Então tem essa discussão, libera na hora ou não. Eu acho que o ideal é liberar na hora, mas aí eu preciso ser remunerada, não é remunerada em dinheiro, mas reconhecida pelo trabalho que fiz criando o dado. Atualmente não se reconhece esse trabalho, em nenhum lugar do mundo, tá?

A proposta de P08, bem como os outros fatores aqui discutidos, é aplicável apenas para seu campo de conhecimento, a medicina. Na biologia, o receio de perder o ineditismo existe, segundo P07.

P07: Há um medo assim por parte de alguns pesquisadores, eu confesso que eu já tive esse medo, principalmente quando a gente tá na fase do mestrado e doutorado, que você tem medo: "não, vão pegar meus dados, vão publicar meus dados".

Entretanto, conforme discutido anteriormente, no campo da biologia as exigências para disponibilização do conjunto de dados em repositórios adequados têm sido apresentadas recorrentemente como condição para a publicação de um artigo. Já no campo da farmácia, o processo parece operar de forma contrária às tendências da ciência aberta. Comparativamente, pesquisas no campo da farmácia mostraram-se menos propícias à abertura do conjunto de dados no momento da publicação ou após um período de exploração exclusiva, uma vez que pode originar novas publicações e, principalmente, novas tecnologias. No caso das tecnologias, as informações resultantes da pesquisa bem como seus dados são comunicados em formato de patente. Conforme discutido anteriormente, isso garante ao autor ou ao detentor dos direitos patrimoniais a exploração exclusiva do seu conteúdo.

P03: O que tá na publicação é o que é pra publicar, porque às vezes o que não está na publicação, esse é segredo de justiça. [...] A gente escolhe assim, na realidade, como a gente trabalha com IT, iniciação tecnológica, a gente escolhe, “isso aqui vai para publicação, isso aqui fica guardado”. Esse que fica guardado a gente não passa e o outro que sintetiza, às vezes ele diz assim: “ah, não, esse aí não, esse é patente”.

A necessidade de manter o sigilo dos dados da pesquisa é tão acentuada no campo da farmácia que até mesmo as práticas de registro dos dados são adequadas a esse objetivo. Segundo o relato de P03, os dados considerados mais promissores pela pesquisadora são registrados em códigos para que outros pesquisadores não sejam capazes de compreender. Portanto, as precauções com o sigilo da pesquisa iniciam-se no processo de produção dos dados e são refletidas nas etapas de compartilhamento e no reuso (ou na sua ausência).

P03: O caderno de zika africano com os meus resultados, o que me interessa ou então tá em código, porque se tiver em código ninguém vai saber o que que é.

De forma geral, o sigilo das pesquisas foi apresentado como um comportamento necessário às práticas científicas, especialmente aquelas que trazem reconhecimento e recompensas aos pesquisadores pelos seus trabalhos. A intensidade com que o sigilo é almejado variou entre as áreas do conhecimento. Igualmente, observou-se a variação entre as áreas do conhecimento em função da quantidade e qualidade das informações disponibilizadas sobre o vírus da zika. Coerentemente, com tudo que já foi exposto até aqui, os pesquisadores no campo da biologia têm considerado que o fluxo de informação sobre o tema tem sido bom e que os outros pesquisadores têm disponibilizado seus resultados e seus dados de forma rápida, conforme concluiu P05.

P05: [...] Eu acho que hoje a divulgação da pesquisa do zika, eu não acho que tá tendo tanta dificuldade de informação. As publicações estão sendo rápidas, as pessoas tão botando os dados pra fora rápido.

Nas outras duas áreas, medicina e farmácia, observou-se que a disponibilização das informações não é tão rápida e aberta. O pesquisador P09, do campo da medicina, afirmou ter dados ainda não disponibilizados sobre o vírus da zika, embora tenha demonstrado algum interesse em fazê-lo. Já a pesquisadora P03, do campo da

farmácia, afirmou ter conhecimento de dados que estão sendo gerados, que não estão sendo disponibilizados e não apresentam um horizonte para sua abertura.

P09: Ontem mesmo eu tive um seminário aqui, um seminário internacional da [nome da instituição] de Enfermagem e eles me convidaram para falar sobre a perspectiva de controle das doenças tangidas por zika. Algumas coisas que têm lá eu já publiquei, mas outras não, porque tem coisas muito recentes.

P03: As pessoas tão segurando muita coisa. Assim, por exemplo, lá na proteoma eu sei que tem gente fazendo zika, e eu sei mais ou menos, porque ele tem confiança em mim. Aí eu não posso falar, mas quer dizer, são muitas coisas que é sigilo, gente, não adianta.

P04: A gente tem aí um buraco. Tem muitas informações sobre zika que a gente sabe que existem grupos pesquisando, mas a informação em si, o quanto aquela pesquisa já avançou os resultados preliminares nós não temos. Então, existem buracos e de repente o que a gente tá testando agora em bancada, pode ser que um outro grupo já tenha feito, já está mais avançado, mas que a gente não sabe.

5.2.7. Categoria “doenças emergentes”

As práticas de disponibilização dos resultados das pesquisas e seus dados são diferenciadas no contexto das doenças emergentes. A possibilidade de criação de exceções vantajosas para a comunicação dos resultados de pesquisa foi um dos motivos que levou a OMS a declarar os problemas epidemiológicos do vírus da zika uma emergência internacional, segundo a reportagem BRASIL SONEGA (2016).

Conforme explicaram Carmo, Penna e Oliveira (2008), a classificação de uma epidemia como emergência pública teve origem em 1992 em proposição do Institute of Medicine, também conhecido como National Academy of Medicine (NAM), dos Estados Unidos da América. O objetivo da classificação é criar um grande alerta internacional para as doenças infecciosas com aumento de incidência entre seres humanos. A definição proposta em 1992 foi sofrendo alterações, até que, em 2005, foi estabelecida no Regulamento Sanitário Internacional (RSI) como doenças que oferecem risco de saúde pública para outros países, as quais, portanto, requerem resposta internacional coordenada (OMS, 2009). As ações coordenadas são executadas no intuito de vigiar a incidência e controlar o vetor de transmissão. Tais ações são determinadas pela Vigilância Global de Saúde em acordo com o RSI e afeta todos Estados-membros da OMS. Com isso, segundo Bruniera-Oliveira et al. (2014), a classificação de uma epidemia como emergência de saúde pública internacional implica quatro ações:

- notificação de todos os casos à OMS;

- implementação de todas as recomendações da OMS;
- desenvolvimento da capacidade nacional e regional de vigilância epidemiológica;
- compartilhamento dos dados em âmbito nacional e internacional em tempo real.

A classificação da epidemia como uma emergência de saúde pública compromete os Estados-membros da OMS a disseminar todos os dados relacionados com a incidência da doença no seu território, bem como os dados necessários ao controle do fenômeno. A obrigatoriedade na divulgação das informações ainda determina que ela deve ser feita em tempo real, afetando assim as práticas e os interesses até então discutidos dos pesquisadores e da cultura acadêmica como um todo. Essa influência foi confirmada pelos pesquisadores das três áreas, embora a ênfase tenha sido empregada especificamente para o fenômeno do vírus da zika, não abrangendo assim todos os temas de emergência de saúde pública internacional.

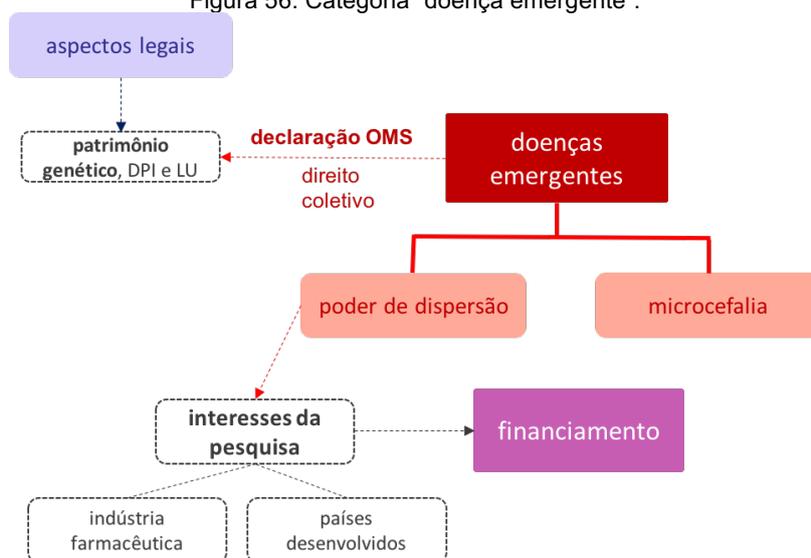
P01: Existe um site que chama “www.virological.org” ou “.com”, acho, onde as pessoas estão divididas em zika, ebola, chikungunya. Então, as pessoas estão fazendo essas análises e eles vão escrevendo o artigo juntos, isso é comum, em virologia, para algumas doenças emergentes, não são todas. Ebola foi uma, chikungunya tá começando, dengue as pessoas... e influenza, principalmente, as pessoas fazem isso com influenza há muito tempo. Porque todo ano tem epidemias novas começando, então essas análises são feitas meio que juntas, assim. Agora, para outras doenças emergentes, bacterianas ou fúngicas, eu acho que isso não é muito comum.

Mesmo na área de farmácia, considerada pelo estudo como a menos propícia para a comunicação aberta dos dados de pesquisa, observou-se que, no contexto específico do fenômeno do vírus da zika, existem comportamentos mais favoráveis à ciência aberta. Segundo o relato de P10, os estudos relacionados ao tema foram os mais progressistas em relação à ampliação da abertura, ao alcance e à rapidez do compartilhamento dos resultados.

P10: Agora com a história do zika, a gente realmente viu uma progressão nesse sentido. Vários trabalhos que foram publicados ao mesmo tempo em que foram submetidos. Então dava para ver que estava submetido, mas já estava disponível. Os dados estavam disponíveis para as pessoas já adiantarem as suas próprias pesquisas. O que é uma coisa de certa forma importante, porque, se você for pensar bem, quanto tempo depois da pesquisa realizada ela é publicada de fato? Muitas vezes isso tem uma demora que é postergada. Então, quando você tem um problema de emergência médica que é o zika, isso pode fazer a diferença.

Os fatores que colocaram o fenômeno do vírus da zika em evidência internacional foi outro elemento investigado com os pesquisadores. De forma geral, o interesse internacional foi atribuído a dois fatores: o poder de dispersão do vírus e a microcefalia em bebês (Figura 56). Portanto, os dois fatores representam a perspectiva dos pesquisadores sobre as razões pelas quais as pesquisas sobre o vírus da zika estão mais propensas a ser comunicadas no contexto da ciência aberta do que outros temas da mesma natureza. Esses resultados respondem ao quarto objetivo específico do estudo. Devido à relevância dos dois fatores para a discussão, ambos serão tratados aqui como subcategorias.

Figura 56: Categoria “doença emergente”.



Fonte: Elaboração própria.

A categoria doenças emergentes representa um fator de interveniência no processo de comunicação dos dados de pesquisa que relaciona-se diretamente como o contexto social da doença. Sua explicação para o fenômeno desdobra-se ainda em outros dois elementos elucidativos do fenômeno, denominados aqui de subcategorias e analisados a seguir.

5.2.7.1. Subcategoria “poder de dispersão do vírus”

As emergências de saúde pública internacional são objeto de interesse e atuação de vários países por meio de protocolos de ações coordenadas. Entretanto nem todas as epidemias sob essa classificação recebem o mesmo interesse das fontes de

financiamento e se destacam nas discussões da ciência dos países desenvolvidos. As condições que fizeram do fenômeno do vírus da zika um tema de interesse e financiamento internacional foram investigadas, a partir da opinião dos pesquisadores. Nesse sentido, a possibilidade de dispersão do vírus para os países ricos teria sido um dos principais motivadores da mobilização internacional.

O pesquisador P10 explica que foi identificado que o vírus tem como uma das suas formas de dispersão a transmissão sexual. Essa forma de transmissão faz com que o vírus possa chegar a novas regiões e não somente àquelas que apresentam as condições de propagação do seu vetor o *Aedes aegypti*. A transmissão sexual do vírus fez com que o tema se tornasse preocupação de vários países, inclusive daqueles mais ricos, atraindo assim mais investimentos e visibilidade para o tema.

P10: O que eu destacaria como principal fator é a transmissão sexual, quando ele foi visto como sendo um agente de transmissão sexual. Tem sido visto que ele tem uma permanência em certas pessoas. A gente tem conhecimento de um caso na USP que o pessoal acompanhou essa pessoa e ele liberava vírus em grandes quantidades pelo sêmen durante seis meses após ter tido os sintomas. Quer dizer, ele é um dispersor de vírus por via sexual. Então isso daí realmente foi o que chamou atenção da comunidade científica dos países mais ricos sobre a questão. Então é essa transmissão sexual que com certeza abalou mais a comunidade.

A transmissão do vírus por vias sexuais ampliou significativamente o seu poder de dispersão. Isso porque outras doenças que acometem grande parte da população de países pobres, e que são igualmente transmitidas pelo mesmo vetor o *Aedes aegypti*, como a dengue e a febre amarela, estão limitadas a regiões de climas tropicais, com pouca infraestrutura de saneamento básico e um deficitário sistema de saúde pública. No entanto, o vírus da zika se mostrou mais “democrático”. A partir do momento em que deixou de depender do seu principal vetor de transmissão, passou a ser uma ameaça global. As preocupações no alcance do vírus atraíram a atenção até mesmo da indústria farmacêutica, considerada negligente com outras doenças da mesma natureza, como a dengue e a febre amarela. O pesquisador P10, da área de farmácia, apontou que o interesse da indústria em investir em pesquisa e tecnologia contra o vírus foi determinado pelo poder lucrativo de tais resultados. Esse poder só foi vislumbrado a partir do momento que ameaçou os países ricos. Portanto, a possibilidade de o vírus da zika alcançar países desenvolvidos atraiu a indústria farmacêutica. No entanto, P10 alerta que esse interesse pode ser enfraquecido e até mesmo desaparecer, caso sua dispersão para os países ricos não se confirme.

P10: A gente não sabe o que vai acabar acontecendo, porque, dependendo da evolução do espalhamento e da dispersão do zika, eles podem perder o interesse. Não é novidade, lógico que eles trabalham atrás de lucro, né? E só vai investir em coisa que seja lucrativa. [...] Na história do zika tem um pouco de mobilização inicial, porque tem esse componente da transmissão sexual que eu te disse que o transforma em um vírus que pode ter uma circulação mais ampla dentro do mundo, atinge todos. Porque a transmissão sexual atinge todos os países, e então aí acaba tendo um pouco do interesse comercial...

5.2.7.2. Subcategoria “microcefalia”

O acometimento das populações pelo vírus da zika pode gerar um conjunto de sintomas prejudiciais à saúde humana. Segundo Vasconcelos (2015), o vírus causa doença febril com a ocorrência de sintomas como mal-estar, edemas e dores nas articulações. Essa doença, apesar de sua aparente benignidade, pode apresentar quadros mais severos, como a síndrome de Guillain-Barré e a ocorrência de microcefalia.

A microcefalia foi consensualmente apontada por todos os pesquisadores entrevistados como a causa da visibilidade internacional do vírus. De acordo com P05, as doenças causadas pelo vírus da zika não são tão severas quando comparadas às outras duas que lhe são associadas, a dengue e a chikungunya, uma vez que essas outras duas doenças têm mais potencial de gerar mortalidade nas populações. Entretanto, a ocorrência da microcefalia, somada ao alto poder de dispersão do vírus, colocou-o no centro das atenções internacionais.

P05: Se você pegar como doença das três, dengue, chikungunya e zika, a zika é a mais fraquinha é a que menos, vamos dizer, é o que a gente chama de morbidade, que tem menos morbidade, vamos dizer assim, mas é o que é terrível por causa da microcefalia. [...] Porque a chikungunya é muito pior pro Brasil. Porque está invalidando as pessoas, tão se tornando inválidas, vão começar a se encostar no INSS à beça, porque as pessoas tão perdendo, por exemplo, a chikungunya leva a lesões articulares importantes que a gente não sabe ainda. Nós temos aí dois anos de acompanhamento, e tem vários pacientes acompanhando em que, por exemplo, enfermeiras que não podem mais trabalhar, não pode mais pegar paciente no meio, não tem força mais, perdeu a força na mão, perdeu a força no ombro, entendeu? O cara que era um pedreiro não pode mais trabalhar em obra, tem um pai de um peruano nosso aqui que também não pode mais trabalhar, tá em casa com licença médica há seis meses. Então, a gente vê um problema social grave também a chikungunya, pior do que a zika, mas a zika tem a microcefalia que é uma coisa que choca, né? É uma coisa. E a dengue mata [...] Se você pegar, a dengue mata muito mais do que o número de casos de microcefalia.

A pesquisadora P08 comparou o fenômeno do vírus da zika ao do ebola, que também foi objeto do acordo de cooperação entre os Estados-membros para a transmissão de informação. Segundo P08, o ebola é mais letal do que as doenças

relacionadas ao vírus da zika, mas estas receberam mais interesse internacional e da ciência por causa do seu poder de dispersão e das ocorrências de microcefalia.

P08: [00:31:01] Ebola por exemplo, que é muito mais letal do ponto de vista, brabo, não despertou assim. Também porque não espalhou pelo mundo com a velocidade que a zika espalhou, né? Mas não despertou a mesma...

MC: E essa diferença seria por conta de se tratar de bebês, da microcefalia?

P08: Da microcefalia, por conta da má formação.

A comoção internacional gerada em torno da ocorrência da microcefalia resultou em importantes investimentos para o combate ao vírus da zika, especialmente em pesquisas sobre o tema. O pesquisador P01 declarou que o vírus em questão já era estudado, mas nunca havia recebido tanta relevância. Sua importância tem sido expressa em grandes financiamentos, o que tem atraído mais interesse de outros pesquisadores para o tema. Portanto, a grande disponibilidade de recursos envolvidos para os estudos sobre o vírus da zika tem determinado o envolvimento de outros pesquisadores com a temática. Segundo P01, outras doenças, como a chikungunya, são mais graves do que aquelas relacionadas ao vírus da zika, mas seu envolvimento com o tema foi justificado pelo acesso aos recursos disponíveis.

P01: Eu vou ser bem sincero. Cientista também é um... como posso dizer isso de maneira educada? As pessoas querem dinheiro, ponto. Primeira coisa. Eu tive um amigo... tive, não, ele ainda é vivo. Ele fez o doutorado estudando zika há três anos atrás, enquanto ele ainda tava lá na África, na banca de defesa do projeto dele e lá ele perguntava: "por que você tá estudando esse tipo de coisa, esse tipo de coisa não importa?" Por quê? Não importa porque uma doença emergente... cinco casos nos últimos vinte anos, não é uma doença relevante. Então, no final das contas, fica um pouco assim: "isso não é importante, a doença nem é uma doença grave". O problema é que associaram com microcefalia. Aí entrou muito recurso, entrou dinheiro, aí agora... entrou dinheiro e qualquer publicação que você fizer, que você tenha feito nos últimos um ano e meio de zika, você publica em revistas que dificilmente a gente conseguiria publicar, por exemplo. Eu vou citar o exemplo do trabalho da Science. Normalmente para você publicar um artigo numa revista assim, com impacto muito alto, você sempre precisa de ter uma quantidade enorme de dados, genoma. Com o zika, eles conseguiram fazer isso com oito genomas, coisa que a gente obtém numa tarde, às vezes, você consegue isso. Então, o dado era muito importante, eles conseguiram isso, mas, se fosse em qualquer outra época, eles não teriam colocado nessa revista, entendeu? Então, no final das contas, a revista vai focar num problema emergencial, porque ela lá vai conseguir muito mais citações se ela colocar esse tipo de artigo agora, o governo tá colocando dinheiro "a rodo"... saiu um edital do CNPq, o resultado saiu semana passada, sessenta milhões. A gente não tem um edital para virologia de sessenta milhões, acho que nunca teve. Então, na verdade, acho que o interesse por zika foi esse, caiu as pessoas... as pessoas estão vendo dengue o tempo todo, chikungunya já tava. Chikungunya é muito mais agressivo, causa muito mais sintoma, deixa a pessoa muito mais debilitada, e chikungunya, até agora, você não viu muita coisa. E é uma doença muito mais importante que zika, porque ela mata mais, inclusive ela mata crianças também, mas ela tem menos visibilidade, ainda, até o momento que ela chegue no status que ela vai ter mais visibilidade. Então, acho que cientistas, eles fazem... e eu incluído, porque só me envolvi com zika porque a gente tinha um

edital da FAP de cinco milhões, quase seis... foi um edital de cinco milhões e pouco. “A gente precisa de recurso no laboratório, vamos trabalhar com zika”. Acaba que eu tô fazendo uma coisa original, mas que eu não faria se não fosse pelo dinheiro. Porque, assim, é uma coisa que tem custo, então eu não usaria dinheiro de um outro projeto para fazer esse tipo de coisa, sabe.

5.3. Teoria substantiva e codificação seletiva

A definição das categorias e subcategorias emergente dos dados das entrevistas é seguida do processo de codificação seletiva. Esse processo representa a elaboração de uma teoria sobre o fenômeno estudado. Sua relevância é explicada por Strauss e Corbin (2008, p. 35): “gerar teorias sobre fenômenos em vez de gerar apenas um conjunto de resultados é importante para o desenvolvimento de um campo de conhecimento”. Os autores pontuam que o seu objetivo é desenvolver teorias úteis a explicação de um fenômeno. No entanto, eles ponderam que há outras formas de pensar sobre teorias e de avaliá-las e que elas variam em função de sua natureza.

Portanto, declara-se aqui que a teoria proposta é um modelo teórico que visa explicar o fenômeno da comunicação dos dados de pesquisa a partir da perspectiva dos pesquisadores brasileiros no campo das ciências biológicas. A discussão do fenômeno dialoga diretamente com as teorias da ciência da informação e, de forma geral, com as ciências sociais.

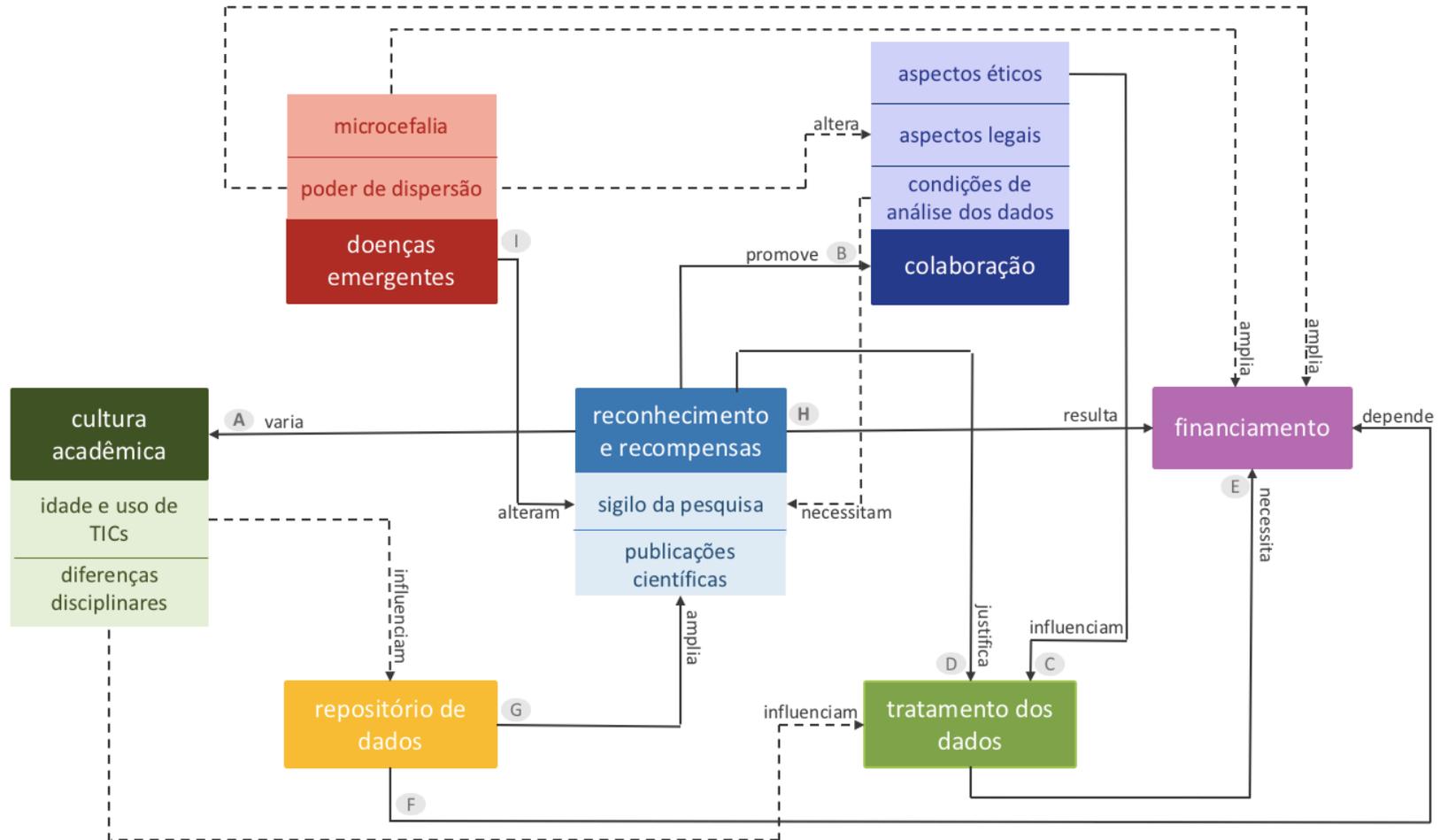
Para a execução da codificação seletiva, segundo a metodologia proposta por Strauss e Corbin (2008), é necessário identificar a categoria central, que se relaciona diretamente com todas as outras categorias e, frequentemente, aparece no conjunto de dados coletados. Para a apresentação da relação, os autores propõem o uso auxiliar de um diagrama. A elaboração da teoria, que consiste no estabelecimento das relações entre as categorias abstratas, é seguida do seu refinamento. Segundo Strauss e Corbin (2008), o refinamento da teoria é a revisão do esquema apresentado, de modo a evidenciar as variações nas categorias e nas suas relações.

5.3.1. Em busca de reconhecimento e recompensa

A categoria “reconhecimento e recompensa” foi a única que se relacionou com todas as outras (Figura 57). Portanto, ela foi considerada como central para a discussão e elemento raiz do processo investigado. A expectativa de serem reconhecidos e recompensados pelos seus trabalhos é a principal motivação para que os pesquisadores divulguem os resultados de suas pesquisas. A inclusão da coleção integral de dados (*dataset*) no conjunto dos produtos resultantes da pesquisa variará em função das diferenças disciplinares.

Capítulo 5 - Fatores que influenciam a comunicação dos dados de pesquisa sobre o vírus da zika na perspectiva dos pesquisadores

Figura 57: Modelo teórico resultante dos resultados da pesquisa



Fonte: Elaboração própria.

Cada uma das relações apresentadas na figura foi descrita a partir de sua letra de identificação. As relações ainda foram qualificadas por meio do uso de verbos indicativos nas setas de referências.

a. A busca de reconhecimento e recompensa em função da cultura acadêmica

As diferenças disciplinares foram analisadas a partir dos tipos de dados e das formas utilizadas para comunicá-los. Essencialmente, foram identificadas três áreas do conhecimento entre os pesquisadores entrevistados, a saber: biologia, medicina e farmácia. Nas pesquisas relacionadas a biologia, foram predominantes os dados oriundos de sequenciamento do genoma, denominados de dados genéticos. Desde o início dos anos 1990, esse tipo de dado é comumente divulgado por meio de repositórios de dados genéticos. As análises realizadas a partir dos dados são comunicadas por meio de artigos de periódicos. Nas pesquisas relacionadas ao campo da farmácia, foram predominantes os dados de volumes quantitativos e imagens da célula. Tais dados são divulgados de forma selecionada em artigos de periódicos e/ou em patentes. Já nas pesquisas relacionadas ao campo da medicina, os dados predominantes relacionam-se com as informações dos pacientes e são exclusivamente apresentados em artigos de periódicos ou outra publicação acadêmica da mesma natureza.

As diferenças disciplinares são um elemento explicativo da cultura acadêmica, ou seja, servem para entender as normas que influenciam o comportamento dos pesquisadores. Além das singularidades das áreas do conhecimento, foram evidenciadas outras duas questões que estão diretamente relacionadas. Os pesquisadores da área de biologia apontaram que, dentro de seu campo de estudo, a idade dos pesquisadores envolvidos influencia o comportamento em relação à disposição para comunicar abertamente seus dados e publicações. Nesse sentido, os pesquisadores mais jovens foram indicados como os mais propensos às práticas, em razão de suas habilidades no uso das tecnologias e de seu interesse em colaborar com outros projetos de pesquisa. Infere-se, portanto, que a biologia é a área do conhecimento que mais utiliza os repositórios de dados, entretanto observam-se variações de comportamento para o seu uso de acordo com a idade dos pesquisadores e suas habilidades com as tecnologias da informação.

Existem formas diferenciadas de comunicar os resultados das pesquisas bem como seus dados entre as áreas do conhecimento. No entanto, os artigos de periódicos

e outras publicações acadêmicas da mesma natureza são comuns a todas as áreas. Em síntese, essas publicações são utilizadas como canal para obtenção de reconhecimento dentro da área acadêmica. Portanto, os periódicos acadêmicos representam um dos elementos explicativos da categoria “reconhecimentos e recompensas”.

O posicionamento dos periódicos, por meio de suas políticas editoriais, variou em função da área do conhecimento. Áreas como a biologia, especialmente sobre o tema da genética, apresentaram como padrão a demanda para a divulgação da coleção dos dados juntamente com a publicação. Na área de medicina, não foi observado um padrão, mas foram identificadas algumas experiências incipientes nesse sentido. Segundo os entrevistados, não existem políticas editoriais na área de farmácia que requeiram a divulgação da coleção dos dados da pesquisa. Nesse campo, apenas parte dos dados são apresentados no artigo de periódico, mantendo-se em sigilo outra parte que possa ser patenteada e/ou gerar tecnologia, como vacinas. Mesmo nas pesquisas voltadas ao desenvolvimento da vacina para o vírus da zika, que é uma doença classificada como emergente e é alvo de financiamento público, existe a expectativa de patentear os resultados.

b. Reconhecimentos e recompensas por meio da colaboração científica

A intenção de obter reconhecimento por meio dos periódicos acadêmicos resultou no segundo elemento explicativo da categoria, a necessidade de manter sigilo da pesquisa. Os pesquisadores pontuaram que é uma exigência dos periódicos que os artigos sejam originais e inéditos, e isso justifica a necessidade de sigilo da pesquisa em razão da demanda dos próprios periódicos. Além disso, justificaram a necessidade em decorrência de suas condições de análise dos dados. Pesquisadores das áreas de biologia e farmácia consideraram que não dispõem de toda a infraestrutura necessária para a análise dos dados com rapidez, tais como os pesquisadores dos países desenvolvidos. Portanto, a comunicação aberta dos dados de pesquisa seria especialmente desvantajosa para os pesquisadores de países em desenvolvimento.

O que torna os países em desenvolvimento prejudicados é a precariedade em relação aos equipamentos e a outros materiais necessários à análise, tais como os reagentes. As singularidades do fenômeno estudado indicam que os pesquisadores brasileiros de fato têm menos acesso aos equipamentos e a outros recursos necessários

às pesquisas, embora tenham mais facilidade de obter amostras biológicas dos pacientes. O discurso dos pesquisadores, bem como os fatos aqui documentados indicam que o material biológico dos pacientes infectados com o vírus tem se tornado um elemento importante para equilibrar as possibilidades de competição no caso dos pesquisadores brasileiros.

O proveito da abundância de material biológico que os pesquisadores brasileiros dispõem ainda contou com o auxílio de um instrumento legal, a Lei nº 13.123/2015. A lei buscou regulamentar o acesso de estrangeiros ao que ela chamou de patrimônio genético. Com isso, o envio de amostras biológicas para o exterior foi dificultado. O fenômeno fez com que pesquisadores de outros países buscassem o acesso ao material biológico por meio da colaboração com pesquisadores brasileiros.

No cenário apresentado, as condições dos pesquisadores brasileiros no que se refere à análise dos dados tornaram-se um elemento importante para justificar a colaboração entre os pesquisadores brasileiros e estrangeiros. Ao mesmo tempo que os pesquisadores brasileiros necessitam de melhores recursos para analisar os dados, os pesquisadores estrangeiros demandam amostras biológicas para gerar os dados. Entretanto, apesar da reciprocidade da dependência, o processo de análise dos dados é mais valorativo para a ciência do que a coleta dos dados. Via de regra, são as análises dos dados que são publicadas em artigos científicos e resultam em reconhecimentos e recompensas aos pesquisadores. Os esforços relacionados com a coleta e o tratamento dos dados por sua vez não geram os reconhecimentos desejados pelos pesquisadores, tornando-se assim apenas um meio para realizar a análise e publicá-la.

c. A colaboração como alternativa para o tratamento dos dados

As práticas de colaboração tornam-se oportunas no contexto descrito. Além das condições necessárias para a análise dos dados e os aspectos legais, outro fator auxilia na elucidação da categoria “colaboração” – os aspectos éticos. O compromisso que os pesquisadores assumem em preservar a privacidade e a dignidade dos participantes envolvidos na pesquisa demanda que os dados sejam divulgados de forma a não identificá-los. A retirada ou descaracterização dos dados pode afetar o entendimento do conjunto dos dados por pesquisadores externos à pesquisa.

O material biológico é instrumento para a geração de dados nas três áreas do conhecimento identificadas, mas é apenas a área da medicina que trata diretamente

com as informações e o contato com os pacientes. Portanto, essa área está mais envolvida nas questões relacionadas à ética na pesquisa do que as demais. Especialmente no campo da medicina, foi sugerido que a disponibilização do conjunto de dados para a utilização por outras pesquisas implica a participação de pelo menos um pesquisador que participou da coleta e análise dos dados em seu processo original.

Os empecilhos para o reuso de dados é consequência de necessidades de esclarecimentos contextuais referentes ao ambiente da pesquisa e principalmente da forma de tratamento dos dados. Portanto o tratamento dos dados foi considerado como outra categoria explicativa do fenômeno, implicada em diferenças disciplinares. Para as pesquisas relacionadas ao campo da medicina, o registro dos dados com vistas a sua divulgação integral implica a padronização dos valores, de forma que sejam compreendidos pelos interessados, apesar de estarem anonimizados.

No campo da farmácia, o principal desafio em relação ao tratamento diz respeito ao registro de toda a coleção de dados em formato digital. Devido às peculiaridades dos laboratórios onde são realizados os experimentos e observações, o uso de computadores é proibido. Com isso, todos os dados são anotados em cadernos e, posteriormente, são transcritos para o formato digital. No entanto, é feita a seleção dos dados que serão transcritos. A seleção contempla apenas os interesses da pesquisa à qual está estritamente relacionada, desconsiderando assim os seus possíveis reusos.

As práticas de tratamento dos dados em pesquisas relacionadas à biologia mostraram-se mais favoráveis à comunicação aberta. Os dados são originalmente registrados em computador, a partir de padrões internacionalmente estabelecidos. O esforço adicional requerido para a comunicação aberta é a adoção do idioma inglês em todo o processo de registro dos dados.

d. O tratamento dos dados em função da obtenção de reconhecimentos e recompensas

Conclui-se que o tratamento dos dados com vistas a sua comunicação aberta demanda mais esforços do que aqueles necessários para o uso particular dos pesquisadores. Embora haja barreiras nesse processo, dois fatores se mostraram como compensadores dos esforços. O primeiro deles relaciona-se com a possibilidade de publicar os resultados da pesquisa em periódicos de grande prestígio internacional. Para

tanto, os pesquisadores tendem a se adequar às políticas editoriais de tais periódicos e alterar a forma de registro dos dados, caso seja necessário.

e. O financiamento do tratamento dos dados

Outra forma de promoção da comunicação aberta se dá por meio do financiamento do tratamento adequado. Recursos destinados ao tratamento dos dados seriam investidos em serviços, com a contratação de especialistas em dados e sistemas de informação qualificados para diminuir os esforços.

f. O financiamento dos repositórios de dados de pesquisa

O financiamento foi indicado como um elemento essencial para viabilizar a comunicação aberta dos dados, seja por meio do tratamento dos dados seja como meio de viabilizar a infraestrutura necessária, onde se destacaram os repositórios de dados. Além de ofertarem serviços de informação, esses repositórios são sistemas de informação responsáveis por armazenar, organizar, disseminar e preservar os dados de pesquisa. Os serviços pressupõem a mediação por profissionais da informação, que, vinculados à instituição, representam o interesse e o compromisso institucional no tratamento e na disponibilização dos dados das pesquisas que financiam. Ao mesmo tempo em que o compromisso é estruturado, as instituições financiadoras de pesquisas podem utilizá-lo como instrumento para controlar os resultados de seus financiamentos. A respeito dessa categoria, é relevante destacar que as discussões acerca dos repositórios foram em termos de idealização, uma vez que a oferta desse tipo de sistema e serviço não foi identificada no contexto institucional de nenhum dos pesquisadores entrevistados. Apesar da ausência da estrutura, os repositórios foram recorrentemente apontados como elementos necessários para operacionalizar a comunicação aberta dos dados das pesquisas.

g. Os repositórios de dados de pesquisa como instrumento para obtenção de reconhecimentos e recompensas

A utilização de repositórios de dados de pesquisa na busca por reconhecimentos e recompensas se dá em dois sentidos. O primeiro deles é no sentido de ampliar o alcance do compartilhamento dos resultados, de forma estruturada. Especialmente em

áreas como a biologia, identificou-se a expectativa de os repositórios diminuírem os esforços necessários para os registros dos dados e, ao mesmo tempo, gerarem mais visibilidade para as pesquisas e para as instituições às quais as pesquisas estão vinculadas. A visibilidade alcançada a partir de tal sistema é uma condição necessária para a obtenção de mais reconhecimento, que posteriormente poderá resultar em recompensas.

O segundo sentido para a utilização dos repositórios de dados se dá em direção à viabilização do financiamento da pesquisa. Portanto, os sistemas poderiam ser instrumentos utilizados para prestação de contas das pesquisas financiadas, além de auxiliarem na avaliação dos resultados obtidos.

h. A busca de reconhecimento e recompensa em função do financiamento

Na perspectiva dos pesquisadores, os periódicos acadêmicos são o principal meio de reconhecimento de seus trabalhos. Tal reconhecimento gera dois resultados almejados: o status acadêmico, que varia em função da cultura acadêmica, e a ampliação de possibilidades para o financiamento das pesquisas. O financiamento foi apresentado como um meio de obtenção da infraestrutura necessária ao desenvolvimento da pesquisa e como resultado compensatório. Portanto, o financiamento é um meio para comunicação aberta dos dados e também representa a finalidade do processo. Com isso, admite-se que as instituições financiadoras de pesquisas são fatores determinantes do processo de comunicação aberta dos dados de pesquisa, bem como os periódicos acadêmicos.

O financiamento das pesquisas envolvidas com a temática do vírus da zika diferencia-se daquele recebido para outras doenças emergentes pela sua abundância de recursos. Os pesquisadores entrevistados relataram que suas pesquisas sobre a temática contam com valores expressivos de financiamento quando se comparam às pesquisas sobre outros temas. O aumento na quantidade de recursos disponíveis foi imputado às instituições brasileiras públicas de financiamento de pesquisa e ao singular interesse dos países ricos sobre a temática.

Esse interesse foi atribuído especialmente ao poder de dispersão do vírus no mundo, fator que se diferencia substancialmente de outras epidemias de doenças emergentes típicas de países em desenvolvimento. A ameaça mundial do vírus foi anunciada quando se constatou a sua possibilidade de transmissão sexual. A partir de

então, a dispersão do vírus deixou de estar condicionada a questões climáticas e sanitárias típicas de países em desenvolvimento. O processo de dispersão mundial do vírus foi atribuído ao Brasil por ocasião da realização da Copa do Mundo FIFA, em 2014, período em que o país recebeu turistas de várias nacionalidades. Após o evento, foram registrados alguns casos de pacientes de países europeus e norte-americanos contaminados pelo vírus.

O poder de dispersão do vírus atraiu novos financiadores de pesquisa para o tema, agências de países desenvolvidos e o interesse da indústria farmacêutica. Portanto, a incidência do vírus em pacientes de países desenvolvidos, além de atrair o financiamento de agências dos próprios países, conquistou ainda a atenção de corporações farmacêuticas. O interesse do grupo para a temática é relevante para a discussão, pois se apresenta como atípico, uma vez que outras doenças da mesma natureza, algumas até mesmo consideradas mais graves do que a zika, não são objeto de interesse da indústria farmacêutica. A dengue, a chikungunya e a febre amarela são doenças transmitidas pelo mesmo vetor da zika, e, segundo os pesquisadores entrevistados, potencialmente resultam em consequências mais graves do que a zika. No entanto, elas não atraíram interesse da indústria farmacêutica, nem a mesma atenção dos países desenvolvidos, apesar de contarem com financiamento de agências governamentais brasileiras.

i. A busca por reconhecimento e recompensa no contexto das doenças emergentes

As investigações acerca do vírus da zika ganharam visibilidade a partir de estudos realizados no Nordeste brasileiro. A sua dimensão internacional foi alcançada a partir da constatação de seu alto poder de dispersão e da notificação de casos de microcefalia em recém-nascidos. Portanto, na perspectiva dos pesquisadores brasileiros, esses dois fatores foram determinantes para a visibilidade internacional do tema.

Os dois aspectos somados ao processo epidêmico causado pelo vírus resultaram na classificação da epidemia como uma emergência internacional de saúde pública. A classificação é feita pela OMS e requer ações conjuntas de todos os seus países-membros afetados. Uma dessas ações diz respeito a divulgação de informações sobre os incidentes, bem como disponibilização de dados, incluindo amostras

biológicas. Na época que antecedeu a classificação da OMS, instituições brasileiras de pesquisa estavam envolvidas em polêmicas com pesquisadores estrangeiros a respeito da impossibilidade de enviar dados para o exterior. Documentos discutidos aqui sugeriram que a classificação foi oportuna para resolver tal impasse.

A dificuldade no envio dos dados, especialmente das amostras biológicas, relacionava-se ao cumprimento da Lei nº 13.123/2015, que passou a tratar as amostras biológicas colhidas no Brasil como patrimônio genético nacional. Tal legislação baseia-se nos termos presentes nos direitos de propriedade intelectual, garantindo assim ao seu autor ou detentor dos direitos autorais exploração exclusiva do material. No entanto, o direito não é absoluto e pode ser revisto no caso de direito coletivo. Portanto, o entendimento das doenças causadas pelo vírus da zika como de emergência internacional é um fator que retira a barreira legal que impede o envio de dados a pesquisadores estrangeiros.

6. Considerações finais

A comunicação dos dados de pesquisa é condição necessária para o cumprimento dos objetivos da ciência aberta. A nova forma de pensar a ciência e planejar seus processos tem por base a filosofia livre, emergente a partir dos anos 1990, bem como a expectativa de adequação aos princípios da ciência moderna e sua função social. Especialmente na área das ciências da saúde, a comunicação dos dados de pesquisa tem por finalidade o aceleração da produção do desenvolvimento científico e tecnológico e, com isso, o melhoramento da saúde humana. Tais expectativas foram declaradas também no início dos anos 1990, a partir do PGH e reafirmadas posteriormente a partir de políticas de informação de importantes instituições envolvidas com a promoção da saúde coletiva.

Recentemente, o fenômeno do vírus da zika e do ebola evidenciou novamente a necessidade de comunicar abertamente os dados e as publicações científicas resultantes de pesquisa sobre os temas. O compromisso foi formalizado no início de 2016. Portanto foram analisados os fatores que influenciaram a discussão e a prática da abertura dos dados de pesquisa sobre a temática.

Fatores que influenciam a comunicação dos dados de pesquisa a partir da literatura

A primeira parte do estudo foi realizada por meio da aplicação da técnica da revisão sistematizada da literatura. Como resultado foi possível a identificação dos fatores que são apontados pela literatura da ciência da informação como intervenientes nos processos de comunicação dos dados de pesquisa. A análise considerou a discussão apresentada por 13 artigos selecionados no corpus da literatura da área. Todos os fatores extraídos da discussão foram sistematizados tendo como referência o modelo conceitual da pesquisa apresentado na metodologia do trabalho. Tal modelo evidenciou os pressupostos do estudo e as dimensões de análises consideradas a partir da fundamentação teórica.

Os artigos analisados apresentaram elementos intervenientes para o processo da comunicação dos dados a partir de três dimensões de análise: os imperativos da ciência moderna, as questões relacionadas à infraestrutura e os aspectos referentes ao financiamento das pesquisas. A quarta dimensão, referente ao contexto social da doença, não foi diretamente explorada pelos artigos em decorrência de suas áreas de

conhecimento originárias. No entanto, embora os artigos tenham sido exclusivamente selecionados no campo da ciência da informação, alguns deles levantaram a relevância do fenômeno nas ciências da saúde, considerando, portanto, que a promoção da saúde pública é um fator que promove a comunicação dos dados de pesquisa.

No que diz respeito aos imperativos da ciência moderna, foram identificados nove fatores intervenientes. A consequência de estímulo ou inibição dependerá da ausência ou presença de tais fatores. São eles:

- aspectos éticos;
- aspectos legais;
- assédio aos pesquisadores;
- aceleração da ciência;
- cultura acadêmica;
- recompensas ao pesquisador;
- reuso dos dados;
- transparência da ciência;
- visibilidade.

As questões relativas ao financiamento das pesquisas também se mostraram relevantes para a análise da influência nos processos de comunicação dos dados de pesquisa. A literatura apontou que as instituições provedoras dos recursos necessários ao desenvolvimento da pesquisa são agentes importantes para a determinação do comportamento e das práticas envolvidas na pesquisa. Portanto, entre os elementos relacionados com o financiamento, a literatura analisada apontou os seguintes fatores intervenientes:

- custos envolvidos nos processos;
- financiamento público da pesquisa e noção dos resultados como um bem público;
- posicionamento da instituição financiadora.

A discussão acerca do financiamento dos recursos necessários para o desenvolvimento da pesquisa envolveu necessariamente aspectos relacionados à infraestrutura. No entanto, observou-se que, para a ciência da informação, os aspectos relacionados à infraestrutura adquirem espaço privilegiado na análise do fenômeno, portanto a infraestrutura foi tratada como uma dimensão de análise. Nessa dimensão, reúnem-se os seguintes fatores intervenientes:

- aspectos acerca da gestão dos dados de pesquisa;

- padrões para interoperabilidade;
- políticas de informação;
- sistemas de informação.

Os fatores intervenientes apontados pela literatura da ciência da informação serviram também aos propósitos da metodologia do estudo aqui proposto. Portanto, o conjunto de fatores aqui elencados orientaram a elaboração do instrumento de coleta de dados da segunda etapa da pesquisa, que visou identificar os fatores intervenientes a partir da perspectiva dos pesquisadores.

Fatores que influenciam a comunicação dos dados de pesquisa a partir da percepção dos pesquisadores

Os pesquisadores entrevistados evidenciaram dez novos fatores intervenientes do processo de comunicação dos dados de pesquisa acerca das doenças causadas pelo vírus da zika. Em síntese, os novos fatores trataram do contexto social da doença, dimensão pouco explorada pela literatura da ciência da informação. Os pesquisadores entrevistados apontaram quatro fatores relacionados diretamente à conjuntura da doença que influencia as práticas de comunicação dos seus dados de pesquisa. São eles:

- a classificação da doença como uma emergência internacional de saúde pública;
- o interesse e o financiamento dos Estados Unidos em pesquisas sobre o tema;
- a incidência da microcefalia em bebês gerados por pacientes infectadas pelo vírus;
- o poder de dispersão do vírus pelo mundo.

Além do contexto social da doença, outros seis fatores foram evidenciados a partir da entrevista dos pesquisadores. Alguns deles já haviam sido tratados pela literatura da ciência da informação; no entanto, na perspectiva dos pesquisadores, posicionaram-se com maior centralidade na explicação do fenômeno estudado. Tais categorias são:

- As condições de análise dos dados e suas implicações nas práticas de colaboração entre países desenvolvidos e países em desenvolvimento.

- A relevância dos reconhecimentos e recompensas como pressuposto motivador para as práticas de pesquisa e de cultura acadêmica.
- A manutenção do ineditismo dos resultados da pesquisa como uma forma de obtenção dos reconhecimentos e recompensas almejados pelos pesquisadores.
- O pressuposto do sigilo da pesquisa como um aspecto da cultura acadêmica e condição necessária para obtenção dos reconhecimentos e recompensas almejados pelos pesquisadores.
- A centralidade dos periódicos científicos na atribuição de reconhecimento e recompensas almejados pelos pesquisadores.
- A indicação de idiomas específicos para o tratamento dos dados como condição para sua interoperabilidade e reuso.

Somados aos dez fatores intervenientes já apresentados estão 19 outros fatores evidenciados pela literatura da ciência da informação. Portanto, na perspectiva dos pesquisadores relacionados à temática identificou-se um conjunto de 29 fatores que exercem influência no contexto estudado.

Fatores convergentes e divergentes nas perspectivas identificadas sobre a comunicação de dados de pesquisa sobre as doenças causadas pelo vírus da zika

Por meio da codificação axial dos dados, foi possível determinar os fatores convergentes e divergentes nas perspectivas identificadas pelo estudo. Os fatores identificados diferenciaram-se substancialmente em função da área do conhecimento na qual os pesquisadores entrevistados realizavam suas pesquisas. Em síntese, os fatores convergentes preponderaram nas seguintes categorias:

- Práticas de colaboração, de onde derivam as subcategorias “aspectos éticos”, “aspectos legais” e “condições de análise dos dados”.
- Influências recebidas pelas questões relacionadas ao financiamento.
- A expectativa e a centralidade conferida à categoria “reconhecimento e recompensas” no processo de desenvolvimento das pesquisas e na comunicação dos dados. A categoria que se mostrou como central para a discussão dos fatores intervenientes ainda derivou de outras duas subcategorias: o pressuposto do “sigilo da pesquisa” e o produto

almejado materializado por meio das publicações acadêmicas, especialmente os periódicos científicos.

- As condições de colaboração implicadas na classificação de uma doença considerada emergência internacional, principalmente em razão dos dois fatores classificados como subcategorias “microcefalia” e “poder de dispersão do vírus”.

Já os fatores divergentes evidenciaram-se especialmente em função das diferenças disciplinares e predominam nas seguintes categorias:

- A cultura acadêmica expressa as normas e os valores que orientam o comportamento dos pesquisadores. No estudo, os aspectos da cultura acadêmica foram explorados sob dois aspectos, classificados como subcategorias. O primeiro deles trata das diferenças disciplinares que foram evidenciadas a partir de dois elementos: os tipos de dados e as formas de comunicá-los. O segundo aspecto diz respeito à idade dos pesquisadores e suas habilidades com o uso de tecnologias da informação.
- O tratamento dos dados é outro fator que varia entre as áreas do conhecimento, principalmente em função dos aspectos identificados na cultura acadêmica.
- O uso de repositórios de dados de pesquisa também é definido pela cultura acadêmica, que determina as práticas de pesquisa e de comunicação. Portanto, foram observadas significativas diferenças entre as áreas do conhecimento no que diz respeito ao uso e ao conhecimento da funcionalidade dos repositórios de dados de pesquisa.

Razões pelas quais os dados de pesquisa sobre o vírus da zika estão mais propensos a serem comunicados no contexto da ciência aberta

As razões pelas quais o tema relacionado ao vírus da zika se mostrou mais propenso às práticas de comunicação aberta foram elucidadas na categoria “doenças emergentes”. Portanto, na perspectiva dos pesquisadores, a classificação das doenças causadas pelo vírus da zika como doenças de emergência internacional de saúde pública conferiu-lhe maior visibilidade, favorecendo-as como objeto de colaboração científica.

A própria classificação atribuída pela OMS é um fator que lhes confere tais condições de colaboração, uma vez que a classificação implica a execução de ações conjuntas entre os países membros da OMS. Entre elas, destaca-se a ampla divulgação de informação sobre a incidência do vírus e dados dos pacientes acometidos pelas suas doenças. A classificação foi especialmente necessária para deslocar os empecilhos oriundos da legislação brasileira sobre o patrimônio genético, que dentre outras implicações dificultava o envio de dados de amostras biológicas para pesquisadores estrangeiros. Portanto, conclui-se que a classificação de uma doença em emergência internacional de saúde pública é um fator que intervém no processo de comunicação, conferindo-lhe maior capacidade de colaboração e visibilidade.

No entanto, mesmo em comparação às outras doenças classificadas como emergentes, o tema do vírus da zika ainda se mostra em condições mais propícias para a comunicação aberta dos dados de pesquisa. Isso porque, segundo a perspectiva dos pesquisadores, o poder de dispersão do vírus ameaçou regiões com melhores recursos para o desenvolvimento científico. Como consequência, pesquisadores dessas regiões, agências financiadoras e até mesmo a indústria farmacêutica passaram a se interessar pelo tema e buscar colaboração com os pesquisadores brasileiros que possuem amplo acesso a determinados dados necessários as pesquisas. Logo, o que diferenciou a zika das outras doenças emergentes, como a dengue e a febre amarela foi a ameaça real do vírus alcançar países ricos, onde circulam mais recursos de financiamento de pesquisa e configura-se em regiões que recebem maior atenção da indústria farmacêutica.

Outro fator considerado relevante para colocar o tema em evidência foi a incidência da microcefalia em bebês gerados por pacientes infectadas pelo vírus. A consequência do vírus no processo de reprodução humana despertou comoção internacional, atraindo assim mais atenção e investimento para o tema. A relação do vírus e a malformação de fetos, somada a ampla possibilidade de alcance do vírus por meio da sua transmissão sexual resultou em mais interesse internacional para o tema e maiores ofertas de financiamento, especialmente quando comparada a outras doenças da mesma natureza.

Modelo teórico representacional dos fatores que exercem influência na comunicação dos dados de pesquisa sobre o vírus da zika

O modelo teórico que representa os fatores que influenciam a comunicação dos dados de pesquisa foi elaborado a partir da estratégia da codificação seletiva, que tem como consequência a proposição de uma teoria substantiva. A teoria, portanto, evidencia as relações entre as categorias e as subcategorias no processo de explicação do fenômeno estudado.

A teoria emergente dos dados evidencia a centralidade da expectativa de reconhecimento e recompensas no processo de comunicação dos dados. Tal expectativa, ponderada pelos aspectos relacionados ao financiamento da pesquisa e da cultura acadêmica, influencia os três macroprocessos da comunicação dos dados de pesquisa. Entre os processos de uso e produção dos dados de pesquisa evidenciam-se os fatores relacionados à colaboração, especialmente aqueles que se referem às condições de análise dos dados. Já entre os processos de produção e compartilhamento dos dados, destacam-se os fatores relacionados ao tratamento dos dados e ao uso de repositórios de dados de pesquisa e sistemas de informação dessa natureza. O compartilhamento também é afetado por fatores relacionados à colaboração, especialmente os aspectos éticos e os aspectos legais relacionados aos resultados da pesquisa. Por fim, o compartilhamento e a possibilidade de reuso dos dados no fenômeno estudado são afetadas diretamente pelo contexto social das doenças emergentes. Tal contexto promove as condições legais, éticas e de infraestrutura necessárias para favorecer a comunicação aberta dos dados.

Contribuições do estudo

O estudo produziu uma teoria substantiva acerca dos fatores que influenciam a comunicação dos dados de pesquisa sobre o vírus da zika. Tais fatores evidenciaram as condições favoráveis e as barreiras para a prática. Portanto, é um instrumento analítico para a compreensão da realidade e permite algumas orientações para intervenção em práticas de comunicação dos dados de pesquisa com vistas à promoção da ciência aberta.

Para o campo da ciência da informação a teoria aqui proposta apresenta-se como uma atualização e complexificação das discussões do campo acerca da comunicação científica. Nesse sentido, a comunicação dos dados de pesquisa é um

fenômeno contemporâneo, necessário para a realização da própria ciência em seu modelo ideal de modernidade, bem como é estratégica para a consolidação dos objetivos da ciência aberta. Todas as dimensões da comunicação científica apresentadas pelo estudo configuram-se como uma preocupação do corpo teórico da ciência da informação. Portanto, a teoria proposta preenche as lacunas de conhecimento do campo, ao mesmo tempo em que enriquece a discussão.

No campo da saúde pública, as contribuições do trabalho orientam-se em uma dimensão prática no sentido de aprimorar os processos de comunicação dos dados de pesquisas relacionadas ao tema do vírus da zika. As contribuições são relevantes para pensar a prática pois fundamentam-se no corpus teórico da ciência da informação e em considerações resultantes de observações empíricas das necessidades e desejos dos pesquisadores vinculados a temáticas da saúde pública. A identificação dos fatores intervenientes significa o reconhecimento das barreiras e facilitadores do processo. Portanto, a partir dos resultados apresentados pelo estudo dispõem-se de um instrumento de auxílio para remoção das barreiras e a ampliação dos facilitadores do processo, de modo que as práticas da ciência aberta no campo das ciências da saúde e especialmente na saúde pública sejam ampliadas e otimizadas.

Sugestão para novos estudos:

O modelo teórico aqui proposto evidencia o papel das instituições financiadoras de pesquisa na comunicação dos dados de pesquisa. No entanto, o tema não foi aprofundado por não se tratar do objetivo desta pesquisa. Portanto, considera-se que a temática é relevante para novos estudos que visem aprofundar a discussão.

A cultura acadêmica, expressada por meio das diferenças disciplinares foi outro elemento que se mostrou de fundamental relevância para o entendimento do fenômeno. Entretanto, assim como as instituições financiadoras, a cultura acadêmica não representou objetivo específico desse estudo e sim foi resultado da própria análise dos dados. Portanto, recomenda-se que estudos futuros considerem tal fator como central para a explicação de fenômenos similares.

A categoria central, denominada reconhecimento e recompensa, representa o fator de interveniência central do fenômeno. Isso significa que tal fator é um elemento-chave para intervenção nos processos de comunicação dos dados. A categoria derivou duas subcategorias, dotadas de importância explicativa. Com isso observa-se que os

periódicos acadêmicos, bem como suas políticas editoriais são elementos centrais para constituição da realidade analisada assim como representa um fator estratégico de mudanças almejadas. Portanto, sugere-se que novos estudos explorem as influências das políticas editoriais na comunicação dos dados de pesquisa.

7. REFERÊNCIAS

ADOLFO, L. **Obras privadas, benefícios coletivos: a dimensão pública do direito autoral na sociedade da informação**. Tese de doutorado—Vale do Rio do Sinos: Universidade do Vale do Rio do Sinos, 2006.

ARAKI, H. et al. GeneSetDB: A comprehensive meta-database, statistical and visualisation framework for gene set analysis. **FEBS Open Bio**, v. 2, n. 1, p. 76–82, 1 jan. 2012.

ARCHAMBAULT, E. **Proportion of Open Access Papers Published in Peer-Reviewed Journals at the European and World Levels—1996–2013**. Disponível em: <<http://sciencematrix.com/en/publications/reports/proportion-of-open-access-papers-published-in-peer-reviewed-journals-at-the>>. Acesso em: 4 dez. 2015.

ATKINS, D.; ET AL. **Revolutionizing science and engineering through cyberinfrastructure: report of the National Science Fundation**. Washington, DC: National Science Fundation, 2003.

BABINI, D. O risco do acesso aberto ficar atrelado às editoras comerciais existentes – a necessidade de um sistema global de publicações acadêmicas de acesso aberto não comercial. **Revista Eletrônica de Comunicação, Informação & Inovação em Saúde**, v. 8, n. 4, 19 dez. 2014.

BARROS, M. **A Primavera Acadêmica e o custo do conhecimento**. Journal article (Online/Unpaginated). Disponível em: <<http://eprints.rclis.org/18279/>>. Acesso em: 30 mar. 2016.

BOND, C. S. et al. The conceptual and practical ethical dilemmas of using health discussion board posts as research data. **Journal of Medical Internet Research**, v. 15, n. 6, p. e112, 2013.

BORGMAN, C. L. **Research Data: Who Will Share What, with Whom, When, and Why?** Rochester, NY: Social Science Research Network, 1 out. 2010. Disponível em: <<http://papers.ssrn.com/abstract=1714427>>. Acesso em: 5 out. 2016.

BORGMAN, C. L. **The Conundrum of Sharing Research Data**. Rochester, NY: Social Science Research Network, 21 jun. 2011. Disponível em: <<https://papers.ssrn.com/abstract=1869155>>. Acesso em: 18 out. 2016.

BORKO, H. Information science: What is it? **American Documentation**, v. 19, 1969.

BOULTON, G. Reinventing Open Science for the 21st Century. In: **Uma década de acesso aberto na UMinho e no Mundo**. Lisboa: Universidade do Minho, 2013. p. 239–250.

BOUTAYEB, A. Developing countries and neglected diseases: challenges and perspectives. **International Journal for Equity in Health**, v. 6, n. 1, p. 1–4, 26 nov. 2007.

BRANCO, S. Brazilian copyright law and how it restricts the efficiency of the human right to education. **Sur. Revista Internacional de Direitos Humanos**, v. 4, n. 6, p. 120–141, 2007.

BRASIL. **Constituição da República Federativa do Brasil de 1988**. Brasília: Senado Federal, 1988.

BRASIL. 13.123. Lei Nº 13.123, de 20 de maio de 2015. . 20 maio 2015 a.

BRASIL, M. DA S. D. DE C. E. Doenças negligenciadas: estratégias do Ministério da Saúde. **Rev Saude Publica**, v. 44, n. 1, p. 200–202, 2010.

COSTA, M. **Fatores que influenciam a comunicação de dados de pesquisa sobre o vírus da zika, na perspectiva de pesquisadores**. 2017. Tese (Doutorado em Ciência da Informação) - Universidade de Brasília, Brasília, 2017.

BRASIL, M. DA S. **Situação Epidemiológica - Dados**. Site do Governo Federal (Brasil). Disponível em: <<http://portalsaude.saude.gov.br/index.php/o-ministerio/principal/leia-mais-o-ministerio/741-secretaria-svs/vigilancia-de-a-a-z/tuberculose/11485-situacao-epidemiologica-dados>>. Acesso em: 26 fev. 2016.

BRASIL, M. DA S. **Protocolo de vigilância e resposta à ocorrência de microcefalia** Brasil: Ministério da Saúde, 2015b. Disponível em: <<http://portalsaude.saude.gov.br/images/pdf/2016/janeiro/22/microcefalia-protocolo-de-vigilancia-e-resposta-v1-3-22jan2016.pdf>>

BRASIL, M. DA S. **Situação Epidemiológica - Dados**. Site do Governo Federal (Brasil). Disponível em: <<http://portalsaude.saude.gov.br/index.php/situacao-epidemiologica-dados-zika>>. Acesso em: 27 fev. 2016c.

BRASIL, M. DA S. Monitoramento dos casos de dengue, febre de chikungunya e febre pelo vírus Zika até a Semana Epidemiológica 03, 2016. **Boletim Epidemiológico**, v. 47, n. 6, 2016a.

BRASIL, M. DA S. **Protocolo de vigilância e resposta à ocorrência de microcefalia: Emergência de Saúde Pública de Importância Nacional - ESPIN** Ministério da Saúde (Brasil), 2016b. Disponível em: <<http://portalsaude.saude.gov.br/images/pdf/2016/janeiro/22/microcefalia-protocolo-de-vigilancia-e-resposta-v1-3-22jan2016.pdf>>. Acesso em: 3 mar. 2016

BRASIL SONEGA. Brasil sonega amostras de zika para pesquisa no exterior, dizem cientistas. **G1 - O Portal de Notícias da Globo**, 4 fev. 2016.

BRUNIERA-OLIVEIRA, R. B.-O. et al. Desenvolvimento da Vigilância Epidemiológica de Fronteira no contexto da Globalização: conceitos e marcos teóricos. **Tempus Actas de Saúde Coletiva**, v. 8, n. 3, p. 75–93, 24 dez. 2014.

BRYANT, J. E. et al. Size heterogeneity in the 3' noncoding region of South American isolates of yellow fever virus. **Journal of Virology**, v. 79, n. 6, p. 3807–3821, mar. 2005.

BRYANT, J. E.; HOLMES, E. C.; BARRETT, A. D. T. Out of Africa: A Molecular Perspective on the Introduction of Yellow Fever Virus into the Americas. **PLoS Pathogens**, v. 3, n. 5, maio 2007.

CANCIAN, N. Governo confirma as três primeiras mortes causadas por chikungunya. **Folha de S. Paulo**, 15 jan. 2016.

CANDELA, L. et al. Data journals: a survey. **Journal of the Association for Information Science and Technology**, v. 66, n. 9, p. 1747–1762, 2015.

CARLSON, S. How to Channel the Data Deluge in Academic Research. **The Chronicle of Higher Education**, v. 54, n. 30, 2008.

CARMO, E. H.; PENNA, G.; OLIVEIRA, W. K. DE. Public health emergencies: concept, characterization, preparation and response. **Estudos Avançados**, v. 22, n. 64, p. 19–32, dez. 2008.

CASTRO, J. F. DE [UNESP. **A relação entre patentes farmacêuticas, doenças negligenciadas e o programa público brasileiro de produção e distribuição de medicamentos**. Dissertação (mestrado)—São Paulo: Universidade Estadual Paulista, Faculdade de Ciências e Letras de Araraquara, 25 abr. 2012.

CHAN, L. et al. **Improving access to research literature in developing countries : challenges and opportunities provided by Open Access**. Trabalho / Paper. Disponível em:

COSTA, M. **Fatores que influenciam a comunicação de dados de pesquisa sobre o vírus da zika, na perspectiva de pesquisadores**. 2017. Tese (Doutorado em Ciência da Informação) - Universidade de Brasília, Brasília, 2017.

<<http://repositorio.unb.br/handle/10482/1009>>. Acesso em: 8 mar. 2016.

CHAVES, M. et al. Dengue, Chikungunya e Zika: a nova realidade brasileira Dengue, Chikungunya and Zika: new brazilian reality. **NewsLab**, v. 132, 2015.

CHILDS, S.; MCLEOD, J. Sharing Research Records and Research Data: Findings from a Research Project in Higher Education. **New Review of Information Networking**, v. 10, n. 2, p. 131–145, 1 nov. 2004.

CHIRAC, P.; TORRELEE, E. Global framework on essential health R&D. **The Lancet**, v. 367, n. 9522, p. 1560–1561, maio 2006.

COMTE-SPONVILLE, A. **Dicionário Filosófico**. [s.l.] Martins Fontes, 2011.

CORTI, L. et al. **Managing and sharing research data: a guide to good practice**. [s.l.] Sage, 2014.

CORTI, L.; ET AL. **Managing and Sharing Research Data: A Guide to Good Practice**. UK: University of Essex, 2009.

COSTA, M. P. DA. **Características e contribuições da via verde para o acesso aberto à informação científica na América Latina**. Dissertação / Dissertation—Brasília: Universidade de Brasília, 2014.

COSTA, S. M. DE S. Filosofia aberta, modelos de negócios e agências de fomento: elementos essenciais a uma discussão sobre o acesso aberto à informação científica. **Ciência da Informação**, v. 35, n. 2, 22 ago. 2006.

COX, A.; VERBAAN, E.; SEN, B. Upskilling Liaison Librarians for Research Data Management. **Ariadne**, n. 70, 2012.

CRESWELL, J. W. **Projeto de pesquisa métodos qualitativo, quantitativo e misto**. [s.l.] Artmed, 2010.

CRESWELL, J. W. **Investigação Qualitativa e Projeto de Pesquisa - 3.ed.: Escolhendo entre Cinco Abordagens**. [s.l.] Penso Editora, 2014.

CROW, R. **PUBLISHING COOPERATIVES: AN ALTERNATIVE FOR SOCIETY PUBLISHERS**. [s.l.] SPARC, 2006. Disponível em: <http://www.sparc.arl.org/sites/default/files/media_files/cooperatives_v1-2.pdf>. Acesso em: 12 fev. 2015.

DALCOLMO, M. P.; ANDRADE, M. K. DE N.; PICON, P. D. Multiresistant tuberculosis in Brazil: history and control. **Revista de Saúde Pública**, v. 41, p. 34–42, set. 2007.

DALLMEIER-TIESSEN, S. et al. Enabling Sharing and Reuse of Scientific Data. **New Review of Information Networking**, v. 19, n. 1, p. 16–43, 2 jan. 2014.

DATA DOCUMENTATION INITIATIVE. **DDI Lifecycle 3.2**. Disponível em: <<http://www.ddialliance.org/Specification/DDI-Lifecycle/3.2/>>. Acesso em: 6 out. 2016.

DATAONE. **DataONE**. Disponível em: <<https://www.dataone.org/>>. Acesso em: 18 out. 2016.

DESCARTES, R. **Discurso sobre o método: para bem dirigir a própria razão e procurar a verdade nas ciências**. Curitiba: Hemus, 2000.

COSTA, M. **Fatores que influenciam a comunicação de dados de pesquisa sobre o vírus da zika, na perspectiva de pesquisadores**. 2017. Tese (Doutorado em Ciência da Informação) - Universidade de Brasília, Brasília, 2017.

- DINIZ, M. H. **Dicionário Jurídico**. 2. ed. ed. São Paulo: Saraiva, 2005.
- DOUGLASS, K. et al. 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**, v. 65, n. 2, p. 251–262, 1 fev. 2014.
- DROESCHER, F. D. As publicações em acesso livre e a avaliação de impacto. **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação**, v. 17, n. 35, 5 dez. 2012.
- DYE, C. et al. Data sharing in public health emergencies: a call to researchers. **Bulletin of the World Health Organization**, v. 2, 2016.
- ECDC, E. C. FOR D. P. AND C. **Mosquito maps**. Organização internacional. Disponível em: <http://ecdc.europa.eu/en/healthtopics/vectors/vector-maps/Pages/VBORNET_maps.aspx>.
- FANIEL, I. M.; JACOBSEN, T. E. Reusing Scientific Data: How Earthquake Engineering Researchers Assess the Reusability of Colleagues' Data. **Computer Supported Cooperative Work (CSCW)**, v. 19, n. 3-4, p. 355–375, 2 set. 2010.
- FOSTER – Facilitate Open Science Training for European Research**. Disponível em: <<https://www.fosteropenscience.eu>>. Acesso em: 7 dez. 2015.
- GARCIA, L. P. et al. **Epidemiologia das doenças negligenciadas no Brasil e gastos federais com medicamentos**. Brasília: IPEA, 2011.
- GARDNER, W. Compelled Disclosure of Scientific Research Data. **The Information Society**, v. 20, n. 2, p. 141–146, 1 abr. 2004.
- GARVEY, W. D. **Communication: The Essence of Science: Facilitating Information Exchange Among Librarians, Scientists, Engineers and Students**. [s.l.] Elsevier, 1979.
- GEZELTER, D. **What, exactly, is Open Science?The OpenScience Project**, 2009. Disponível em: <<http://www.openscience.org/blog/?p=269>>. Acesso em: 4 out. 2016
- GRANT, M. J.; BOOTH, A. A typology of reviews: an analysis of 14 review types and associated methodologies. **Health Information and Libraries Journal**, v. 26, n. 2, p. 91–108, jun. 2009.
- GRAY, J. Jim Gray on eScience: A transformed scientific method. In: **The fourth paradigm: Data-intensive scientific discovery**. [s.l.] Microsoft Research, 2009.
- GUIBAULT, L.; WIEBE, A. **Safe to be open. Study on the protection of research data and recommendation for access and usage**. [s.l.] Universitätsverlag Göttingen, 2013.
- Guidelines on open access to scientific publications and research data in Horizon 2020**. European Commission, , 2013. Disponível em: <<https://www.openaire.eu/guidelines-on-open-access-to-scientific-publications-and-research-data-in-horizon-2020>>
- H2020 General Model Grant Agreement**. European Commission, , 2015. Disponível em: <http://ec.europa.eu/research/participants/data/ref/h2020/mga/gga/h2020-mga-gga-mono_en.pdf>
- HAILEY, D. Scientific harassment by pharmaceutical companies: time to stop. **CMAJ: Canadian Medical Association Journal**, v. 162, n. 2, p. 212–213, 25 jan. 2000.
- HARVIE, D. et al. Publisher, be damned! From price gouging to the open road. **Prometheus**, v. 31, n. 3, p. 229–239, 1 set. 2013.
- COSTA, M. **Fatores que influenciam a comunicação de dados de pesquisa sobre o vírus da zika, na perspectiva de pesquisadores**. 2017. Tese (Doutorado em Ciência da Informação) - Universidade de Brasília, Brasília, 2017.

- HEY, T.; TREFETHEN, A. E-Science, cyberinfrastructure and scholarly communication. In: OLSON, G.; ZIMMERMAN, A.; BOS, N. (Eds.). . **Scientific collaboration on the internet**. Londres: MIT Press, 2008. p. 15–31.
- HIGMAN, R.; EMMOTT, S. Research data management and openness: The role of data sharing in developing institutional policies and practices. **Program**, v. 49, n. 4, p. 364–381, 1 set. 2015.
- HOTEZ, P. J. et al. Incorporating a Rapid-Impact Package for Neglected Tropical Diseases with Programs for HIV/AIDS, Tuberculosis, and Malaria. **PLoS Med**, v. 3, n. 5, p. e102, 31 jan. 2006.
- HRYNASZKIEWICZ, I.; COCKERILL, M. J. Open by default: a proposed copyright license and waiver agreement for open access research and data in peer-reviewed journals. **BMC Research Notes**, v. 5, p. 494, 2012.
- INFOPÉDIA. **Definição ou significado de encriptar no Dicionário Infopédia da Língua Portuguesa com Acordo Ortográfico**. Academic site. Disponível em: <<https://www.infopedia.pt/dicionarios/lingua-portuguesa/encriptar>>. Acesso em: 28 dez. 2016.
- JACOMY, M. et al. ForceAtlas2, a Continuous Graph Layout Algorithm for Handy Network Visualization Designed for the Gephi Software. **PLOS ONE**, v. 9, n. 6, p. e98679, 10 jun. 2014.
- JUNGMANN, D. **A caminho da inovação: proteção e negócios com bens de propriedade intelectual: guia para o empresário**. Brasília: IEL, 2010.
- KÄMPF, C. Acesso aberto ao conhecimento científico tem apoio crescente de cientistas. **ComCiência**, n. 139, p. 0–0, jun. 2012.
- KNIGHT, G. Building a research data management service for the London school of hygiene & tropical medicine. **Program**, v. 49, n. 4, p. 424–439, 1 set. 2015.
- KRITSKI, A. L. et al. Duas décadas de pesquisa em tuberculose no Brasil: estado da arte das publicações científicas. **Revista de Saúde Pública**, v. 41, set. 2007.
- KUHN, T. **A estrutura das revoluções científicas**. 9. ed. São Paulo: Perspectiva, 2005.
- LEITE, F. C. L. **Modelo genérico de gestão da informação científica para instituições de pesquisa na perspectiva da comunicação científica e do acesso aberto**. Tese / Thesis. Disponível em: <<http://repositorio.unb.br/handle/10482/9753>>. Acesso em: 7 set. 2016.
- LOPEZ, F. **Workshop de datos científicos. Introducción**. In: CONICET. , 2015. Disponível em: <<http://pt.slideshare.net/fernandoariellopez/workshop-de-datos-cientficos-introduccion>>. Acesso em: 7 dez. 2015
- LOSOFF, B. Electronic Scientific Data & Literature Aggregation: A Review for Librarians. **Issues in Science and Technology Librarianship**, v. 59, 2009.
- LYNCH, C. A. Institutional Repositories: Essential Infrastructure For Scholarship In The Digital Age. **ARL: A Bimonthly Report**, v. 226, p. 327–336, 2003.
- MEADOWS, A. **A comunicação científica**. Brasília: Briquet Lemos, 1999.
- MEDEIROS, J. DA S.; CAREGNATO, S. E. Compartilhamento de dados e e-Science: explorando um novo conceito para a comunicação científica | Data sharing and e-Science: exploring a new concept for scientific communication. **Liinc em Revista**, v. 8, n. 2, 18 dez. 2012.
- MELERO, R.; HERNÁNDEZ-SAN-MIGUEL, J. Acceso abierto a los datos de investigación, una
- COSTA, M. **Fatores que influenciam a comunicação de dados de pesquisa sobre o vírus da zika, na perspectiva de pesquisadores**. 2017. Tese (Doutorado em Ciência da Informação) - Universidade de Brasília, Brasília, 2017.

- vía hacia la colaboración científica. **Revista española de Documentación Científica**, v. 37, n. 4, 2014.
- MIZUKAMI, P. N. Função social da propriedade intelectual: compartilhamento de arquivos e direitos autorais na CF/88. 29 jun. 2007.
- MOLEDO, L.; OLSZEVIKI, N. **Historia de las ideas científicas: de Tales de Mileto a la máquina de dios**. 3. ed. Ciudad Autónoma de Buenos Aires: Planeta, 2015.
- MOLLOY, J. C. The Open Knowledge Foundation: Open Data Means Better Science. **PLoS Biol**, v. 9, n. 12, p. e1001195, 6 dez. 2011.
- MORAN, M.; ET AL. Neglected Disease Research and Development: How Much Are We Really Spending? - ProQuest. **PLoS Medicine**, v. 6, n. 2, 2009.
- MOREL, C. M. **A pesquisa em saúde e os objetivos do milênio: desafios e oportunidades globais, soluções e políticas nacionais**. Disponível em: <<http://www.redalyc.org/articulo.oa?id=63042999002>>. Acesso em: 24 fev. 2016.
- MORRISON, H. **Freedom for scholarship in the internet age**. Tese de doutorado—Canadá: Simon Fraser University, 2012.
- MORRISON, H. **The Imaginary Journal of Poetic Economics: Dramatic Growth of Open Access 2015 first quarter****The Imaginary Journal of Poetic Economics**, 7 abr. 2015. Disponível em: <<http://poeticeconomics.blogspot.com.br/2015/04/dramatic-growth-of-open-access-2015.html>>. Acesso em: 4 dez. 2015
- MUTEBI, J.-P.; BARRETT, A. D. T. The epidemiology of yellow fever in Africa. **Microbes and Infection / Institut Pasteur**, v. 4, n. 14, p. 1459–1468, nov. 2002.
- NASCIMENTO, É. S.; MAIA-HERZOG, M.; SABROZA, P. C. O acesso público da informação em saúde para as doenças negligenciadas no Brasil. **RECIIS**, v. 6, n. 2, 29 jun. 2012.
- NATIONAL INSTITUTES OF HEALTH. **NIH Data Sharing Policy and Implementation Guidance**Estados Unidos. National Institutes of Health (NIH), , 3 maio 2003. Disponível em: <http://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm#goals>
- NATIONAL INSTITUTES OF HEALTH, N. (EUA). **Our Mission - NCBI**. Disponível em: <<https://www.ncbi.nlm.nih.gov/home/about/mission.shtml>>. Acesso em: 5 jan. 2017.
- NATIONAL SCIENCE FOUNDATION (NSF). **Proposal & Award Policies & Procedures Guide (PAPPG)**NSF, , 2014. Disponível em: <<https://www.nsf.gov/pubs/policydocs/pappguide/nsf15001/index.jsp>>. Acesso em: 7 out. 2016
- OBAMA, B. **Memorandum for the Heads of Executive Departments and Agencies**White House of US, , 2009. Disponível em: <<https://www.whitehouse.gov/node/698>>. Acesso em: 18 out. 2016
- OECD. **Declaration on Access to Research Data from Public Funding**Organization for Economic Co-operation and Development (OECD), , 2004. Disponível em: <http://acts.oecd.org/Instruments/ShowInstrumentView.aspx?InstrumentID=157#_ftn0>
- OECD. **OECD Principles and Guidelines for Access to Research Data from Public Funding**, 2007. Disponível em: <<http://www.oecd.org/dataoecd/9/61/38500813.pdf>>
- OMS, O. M. DE LA S. **Regulamento Sanitário Internacional**Agência Nacional de Vigilância
- COSTA, M. **Fatores que influenciam a comunicação de dados de pesquisa sobre o vírus da zika, na perspectiva de pesquisadores**. 2017. Tese (Doutorado em Ciência da Informação) - Universidade de Brasília, Brasília, 2017.

- Sanitári, , 2009. Disponível em: <http://www.anvisa.gov.br/hotsite/viajante/Regulamento_Sanitario_Internacional_vers%C3%A3o%20para%20impress%C3%A3o.pdf>. Acesso em: 9 jan. 2017
- OMS, O. M. DE LA S. **OMS | Informe mundial sobre la tuberculosis**WHO, , 2014. Disponível em: <http://www.who.int/tb/publications/global_report/es/>. Acesso em: 24 fev. 2016
- PADILLA NAVARRO, P. A. et al. Acceso a datos de investigación e información científica en Chile. **Revista española de Documentación Científica**, v. 36, n. 3, 2013.
- PAMPEL, H. et al. Making Research Data Repositories Visible: The re3data.org Registry. **PLOS ONE**, v. 8, n. 11, p. e78080, 4 nov. 2013.
- PERRY. Archiving of publicly funded research data: A survey of Canadian researchers. **Government Information Quarterly**, v. 25, n. 1, p. 133–148, 2008.
- PIERRO, B. Impasse burocrático desacelera pesquisas com material genético. **Revista Pesquisa Fapesp**, 23 mar. 2016.
- PINTO, R. A função social da propriedade intelectual. **Âmbito Jurídico**, v. 63, 2009.
- PIWOWAR, H. A.; CHAPMAN, W. W. A review of journal policies for sharing research data. **Nature Precedings**, n. 713, 20 mar. 2008.
- PIWOWAR, H. A.; VISION, T. J. Data reuse and the open data citation advantage. **PeerJ**, v. 1, p. e175, 1 out. 2013.
- PONTES, F. Doenças negligenciadas ainda matam 1 milhão por ano no mundo. **Inovação em pauta**, p. 69–73, 2009.
- POPPER, K. **A lógica da pesquisa científica**. 2. ed. São Paulo: Cultrix, 2013.
- RAND, R. The Global Change Data and Information System-Assisted Search for Knowledge (GC-ASK) Project. **D-Lib Magazine**, 1995.
- RAY, J. (ED. . **Research data management: practical strategies for information professionals**. [s.l.] Purdue University, 2014.
- RESEARCH COUNCILS UK. **RCUK Common Principles on Data Policy**. Disponível em: <<http://www.rcuk.ac.uk/research/datapolicy/>>. Acesso em: 18 out. 2016.
- RIBEIRO, D. **O Povo Brasileiro: a formação e o sentido do Brasil**. 2. ed. São Paulo: Companhia das Letras, 1995.
- RODRIGUES, E. et al. **Os repositórios de dados científicos: estado da arte**. [s.l.: s.n.]. Disponível em: <<http://repositorium.sdum.uminho.pt/handle/1822/10830>>. Acesso em: 17 abr. 2015.
- RUTTENBERG, S.; RISHBETH, H. Sir Granville Beynon: A Celebration of his Eightieth Birthday on 24 May 1994 World Data Centres — past, present and future. **Journal of Atmospheric and Terrestrial Physics**, v. 56, n. 7, p. 865–870, 1 maio 1994.
- RUUSALEPP, R. A comparative study of international approaches to enabling the sharing of research data. 2008.
- SANKOH, O.; JJSSELMUIDEN, C. Sharing research data to improve public health: a perspective
- COSTA, M. **Fatores que influenciam a comunicação de dados de pesquisa sobre o vírus da zika, na perspectiva de pesquisadores**. 2017. Tese (Doutorado em Ciência da Informação) - Universidade de Brasília, Brasília, 2017.

from the global south. **The Lancet**, v. 378, n. 9789, p. 401–402, jul. 2011.

SARMENTO E SOUZA, M. F. et al. **Algumas considerações sobre as principais declarações que suportam o movimento Acesso Livre**. 2005 Disponível em: <<http://eprints.rclis.org/8512/>>. Acesso em: 30 mar. 2016

SAYÃO, L. F.; SALES, L. F. Dados de pesquisa: contribuição para o estabelecimento de um modelo de curadoria digital para o país. **Pesquisa Brasileira em Ciência da Informação e Biblioteconomia**, v. 8, n. 2, 29 mar. 2014.

SOARES, G. A. D. O Portal de Periódicos da Capes: dados e pensamentos. **Revista Brasileira de Pós-Graduação**, v. 1, n. 1, 3 set. 2011.

STRAUSS, A.; CORBIN, J. **Pesquisa qualitativa: técnicas e procedimentos para o desenvolvimento de teoria fundamentada**. [s.l.] Artmed, 2008.

STURGES, P. et al. Research data sharing: Developing a stakeholder-driven model for journal policies. **Journal of the Association for Information Science and Technology**, p. n/a–n/a, 1 maio 2015.

TAYLOR, M. **The obscene profits of commercial scholarly publishers** **Sauropod Vertebra Picture of the Week**, 13 jan. 2012. Disponível em: <<http://svpow.com/2012/01/13/the-obscene-profits-of-commercial-scholarly-publishers/>>. Acesso em: 4 dez. 2015

TEIXEIRA, G. Tuberculose na América do Sul: a posição do Brasil. **Boletim de Pneumologia Sanitária**, v. 14, n. 3, p. 133–134, dez. 2006.

TEIXEIRA, M. DA G. et al. Epidemiology of dengue in Salvador-Bahia, 1995-1999. **Revista da Sociedade Brasileira de Medicina Tropical**, v. 34, n. 3, p. 269–274, jun. 2001.

THOMPSON, C. A.; ROBERTSON, W. D.; GREENBERG, J. Where Have All the Scientific Data Gone? LIS Perspective on the Data-At-Risk Predicament. **College & Research Libraries**, p. crl13–500, 2015.

TONET, I. **Método científico: uma abordagem ontológica**. São Paulo: Instituto Lukács, 2013.

TRELEAVEN, P.; EMMOTT, S. Intelligent Media. In: AARTS, E.; ENCARNAÇÃO, P. D.-I. D. H C MUL J. (Eds.). **True Visions**. [s.l.] Springer Berlin Heidelberg, 2006. p. 307–319.

TRINE FJORDBACK SØNDERGAARD; JACK ANDERSEN; BIRGER HJØRLAND. Documents and the communication of scientific and scholarly information: Revising and updating the UNISIST model. **Journal of Documentation**, v. 59, n. 3, p. 278–320, 1 jun. 2003.

TROUILLER, P.; OLLIARO, P. L. Drug development output from 1975 to 1996: what proportion for tropical diseases? **International journal of infectious diseases: IJID: official publication of the International Society for Infectious Diseases**, v. 3, n. 2, p. 61–63, 1999 Winter 1998.

VAN DE SOMPEL, H.; LAGOZE, C. The Santa Fe Convention of the Open Archives Initiative. **D-Lib Magazine**, v. 6, n. 2, fev. 2000.

VASCONCELOS, P. F. C. et al. Genetic Divergence and Dispersal of Yellow Fever Virus, Brazil. **Emerging Infectious Diseases**, v. 10, n. 9, p. 1578–1584, set. 2004.

VASCONCELOS, P. F. DA C. Doença pelo vírus Zika: um novo problema emergente nas Américas? **Revista Pan-Amazônica de Saúde**, v. 6, n. 2, p. 9–10, jun. 2015.

COSTA, M. **Fatores que influenciam a comunicação de dados de pesquisa sobre o vírus da zika, na perspectiva de pesquisadores**. 2017. Tese (Doutorado em Ciência da Informação) - Universidade de Brasília, Brasília, 2017.

VIEIRA, D. E. O.; GOMES, M. Adverse effects of tuberculosis treatment: experience at an outpatient clinic of a teaching hospital in the city of São Paulo, Brazil. **Jornal Brasileiro de Pneumologia**, v. 34, n. 12, p. 1049–1055, dez. 2008.

WALPORT, M.; BREST, P. Sharing research data to improve public health. **The Lancet**, v. 377, n. 9765, p. 537–539, 18 fev. 2011.

WEITZEL, S. R. **As novas configurações do Acesso Aberto: desafios e propostas**. Journal article (Print/Paginated). Disponível em: <<http://eprints.rclis.org/23465/>>. Acesso em: 19 out. 2015.

WELLCOME TRUST. **Sharing research data to improve public health: full joint statement by funders of health research**. Disponível em: <<https://wellcome.ac.uk/what-we-do/our-work/sharing-research-data-improve-public-health-full-joint-statement-funders-health#development-of-the-joint-statement-of-purpose>>. Acesso em: 15 out. 2016.

WHITMIRE, A. L.; BOOCK, M.; SUTTON, S. C. Variability in academic research data management practices: Implications for data services development from a faculty survey. **Program**, v. 49, n. 4, p. 382–407, 1 set. 2015.

WHO, W. H. O. **Zika situation report**. Organização internacional. Disponível em: <<http://www.who.int/emergencies/zika-virus/situation-report/19-february-2016/en/>>. Acesso em: 28 fev. 2016a.

WHO, W. H. O. **Latest Zika situation report** WHO, , 19 fev. 2016b. Disponível em: <http://www.paho.org/hq/index.php?option=com_content&view=article&id=11669&Itemid=41716&lang=en>. Acesso em: 27 fev. 2016

WORLD, T. O. (WTO). **Agreement on Trade-Related Aspects of Intellectual Property Rights** Trade Organization (WTO), , 1994. Disponível em: <https://www.wto.org/english/tratop_e/trips_e/t_agm2_e.htm>. Acesso em: 6 jan. 2017

ZANLUCA, C. et al. First report of autochthonous transmission of Zika virus in Brazil. **Memórias do Instituto Oswaldo Cruz**, v. 110, n. 4, p. 569–572, jun. 2015.

ZATZ, M. Projeto genoma humano e ética. **São Paulo em Perspectiva**, v. 14, n. 3, p. 47–52, jul. 2000.

COSTA, M. **Fatores que influenciam a comunicação de dados de pesquisa sobre o vírus da zika, na perspectiva de pesquisadores**. 2017. Tese (Doutorado em Ciência da Informação) - Universidade de Brasília, Brasília, 2017.

Apêndices

Apêndice A: Codificação dos dados da RSL e categorias preliminares

Autores	Codificação dos dados	Categoria preliminar	Dimensão
Knight (2015)	Contexto das pesquisas em saúde pública	Saúde pública	Contexto social da doença
Melero e Hernández-San-Miguel (2014)	Ciências da saúde	Saúde pública	Contexto social da doença
Melero e Hernández-San-Miguel (2014)	Progresso da saúde pública	Saúde pública	Contexto social da doença
Dallmeier-Tiessen et al. (2014)	Benefícios sociais	Bem público	Financiamento
Dallmeier-Tiessen et al. (2014)	Entendimento público sobre as pesquisas	Bem público	Financiamento
Higman e Emmott (2015)	Recursos públicos e bens públicos	Bem público	Financiamento
Douglass et al. (2014)	Democracia	Bem público	Financiamento
Douglass et al. (2014)	Interesse em tornar os dados públicos	Bem público	Financiamento
Melero e Hernández-San-Miguel (2014)	Custo (tempo e dinheiro) para coleta e análise dos dados	Custo	Financiamento
Higman e Emmott (2015)	Altos custos	Custo	Financiamento
Dallmeier-Tiessen et al. (2014)	Controle dos recursos	Custo	Financiamento
Melero e Hernández-San-Miguel (2014)	Aproveitamento de recursos	Custo	Financiamento
Dallmeier-Tiessen et al. (2014)	Custo e permanência da infraestrutura	Custo	Financiamento
Dallmeier-Tiessen et al. (2014)	Requisito para financiamento de pesquisa	Financiadores	Financiamento
Dallmeier-Tiessen et al. (2014)	Crescimento de financiamento	Financiadores	Financiamento
Douglass et al. (2014)	Financiamento	Financiadores	Financiamento
Douglass et al. (2014)	Políticas mandatórias das financiadoras	Financiadores	Financiamento
Melero e Hernández-San-Miguel (2014)	Uso justo (coleta e análise)	Custo	Financiamento
Melero e Hernández-San-Miguel (2014)	Maximização de conhecimentos	Aceleração da ciência	Imperativos da ciência
Melero e Hernández-San-Miguel (2014)	Maximização do uso dos dados	Aceleração da ciência	Imperativos da ciência
Dallmeier-Tiessen et al. (2014)	Redução do tempo e custo para produção de novos resultados	Aceleração da ciência	Imperativos da ciência
Dallmeier-Tiessen et al. (2014)	Privacidade dos participantes	Aspectos éticos	Imperativos da ciência
Dallmeier-Tiessen et al. (2014)	Custo e padrão para anonimizar os dados	Aspectos éticos	Imperativos da ciência
Douglass et al. (2014)	Direitos para disponibilizar os dados	Aspectos éticos	Imperativos da ciência
Melero e Hernández-San-Miguel (2014)	Privacidade dos participantes	Aspectos éticos	Imperativos da ciência
Melero e Hernández-San-Miguel (2014)	Anonimato dos dados	Aspectos éticos	Imperativos da ciência
Bond et al. (2013)	Permissão para o uso dos dados	Aspectos éticos	Imperativos da ciência
Childs e Mcleod (2004)	Confidencialidade dos dados	Aspectos éticos	Imperativos da ciência
Thompson, Robertson e Greenberg (2015)	Privacidade dos participantes	Aspectos éticos	Imperativos da ciência
Knight (2015)	Permissão dos participantes da pesquisa	Aspectos éticos	Imperativos da ciência

Apêndices

Childs e Mcleod (2004)	Implicações legais	Aspectos legais	Imperativos da ciência
Dallmeier-Tiessen et al. (2014)	Legislação	Aspectos legais	Imperativos da ciência
Sturges et al. (2015)	Licenças de uso	Aspectos legais	Imperativos da ciência
Gardner (2004)	Clareza das legislações	Aspectos legais	Imperativos da ciência
Dallmeier-Tiessen et al. (2014)	Direito de acesso à informação	Aspectos legais	Imperativos da ciência
Dallmeier-Tiessen et al. (2014)	Interesses comerciais	Aspectos legais	Imperativos da ciência
Dallmeier-Tiessen et al. (2014)	Direito de propriedade intelectual	Aspectos legais	Imperativos da ciência
Thompson, Robertson e Greenberg (2015)	Proteção dos direitos de propriedade intelectual	Aspectos legais	Imperativos da ciência
Gardner (2004)	Assédio aos pesquisadores	Assédio aos pesquisadores	Imperativos da ciência
Higman e Emmott (2015)	Área do conhecimento	Cultura acadêmica	Imperativos da ciência
Dallmeier-Tiessen et al. (2014)	Cultura acadêmica	Cultura acadêmica	Imperativos da ciência
Dallmeier-Tiessen et al. (2014)	Idade dos pesquisadores	Cultura acadêmica	Imperativos da ciência
Dallmeier-Tiessen et al. (2014)	Interação com outras disciplinas	Cultura acadêmica	Imperativos da ciência
Dallmeier-Tiessen et al. (2014)	Interdisciplinaridade dos dados	Cultura acadêmica	Imperativos da ciência
Losoff (2009)	Recompensa aos autores pela divulgação dos dados	Recompensa ao pesquisador	Imperativos da ciência
Dallmeier-Tiessen et al. (2014)	Status e promoção da carreira de pesquisador	Recompensa ao pesquisador	Imperativos da ciência
Dallmeier-Tiessen et al. (2014)	Valor acadêmico	Recompensa ao pesquisador	Imperativos da ciência
Melero e Hernández-San-Miguel (2014)	Incentivo	Recompensa ao pesquisador	Imperativos da ciência
Melero e Hernández-San-Miguel (2014)	Interesse dos pesquisadores	Recompensa ao pesquisador	Imperativos da ciência
Childs e Mcleod (2004)	Propósito do reuso dos dados	Reuso	Imperativos da ciência
Thompson, Robertson e Greenberg (2015)	Tempo para tornar os dados utilizáveis por outras pesquisa	Reuso	Imperativos da ciência
Dallmeier-Tiessen et al. (2014)	Reutilização de dados	Reuso	Imperativos da ciência
Dallmeier-Tiessen et al. (2014)	Geração de novos resultados	Reuso	Imperativos da ciência
Dallmeier-Tiessen et al. (2014)	Esforço e tempo para preparar os dados para o reuso	Reuso	Imperativos da ciência
Dallmeier-Tiessen et al. (2014)	Desqualificação dos dados (receio dos pesquisadores)	Reuso	Imperativos da ciência
Dallmeier-Tiessen et al. (2014)	Divergência de análises (receio dos pesquisadores)	Reuso	Imperativos da ciência
Higman e Emmott (2015)	Reuso dos dados	Reuso	Imperativos da ciência
Higman e Emmott (2015)	Aplicabilidade do reuso	Reuso	Imperativos da ciência
Douglass et al. (2014)	Tempo para tornar os dados reutilizáveis	Reuso	Imperativos da ciência
Douglass et al. (2014)	Aplicabilidade do reuso dos dados	Reuso	Imperativos da ciência
Dallmeier-Tiessen et al. (2014)	Integridade da ciência	Transparência da ciência	Imperativos da ciência

COSTA, M. **Fatores que influenciam a comunicação de dados de pesquisa sobre o vírus da zika, na perspectiva de pesquisadores.** 2017. Tese (Doutorado em Ciência da Informação) - Universidade de Brasília, Brasília, 2017.

Apêndices

Sturges et al. (2015)	Transparência da ciência	Transparência da ciência	Imperativos da ciência
Higman e Emmott (2015)	Transparência da ciência	Transparência da ciência	Imperativos da ciência
Gardner (2004)	Compreensão da pesquisa	Transparência da ciência	Imperativos da ciência
Melero e Hernández-San-Miguel (2014)	Qualidade da ciência	Transparência da ciência	Imperativos da ciência
Dallmeier-Tiessen et al. (2014)	Validação dos resultados	Transparência da ciência	Imperativos da ciência
Dallmeier-Tiessen et al. (2014)	Teste de teorias	Transparência da ciência	Imperativos da ciência
Dallmeier-Tiessen et al. (2014)	Reconhecimento da produção	Visibilidade	Imperativos da ciência
Dallmeier-Tiessen et al. (2014)	Ampliação da visibilidade (citação)	Visibilidade	Imperativos da ciência
Higman e Emmott (2015)	Visibilidade da pesquisa	Visibilidade	Imperativos da ciência
Cox, Verbaan e Sen (2012)	Gestão dos dados de pesquisa	Gestão dos dados de pesquisa	Infraestrutura
Whitmire, Boock e Sutton (2015)	Responsabilidades com a gestão dos dados	Gestão dos dados de pesquisa	Infraestrutura
Whitmire, Boock e Sutton (2015)	Serviço de gestão dos dados	Gestão dos dados de pesquisa	Infraestrutura
Thompson, Robertson e Greenberg (2015)	Responsáveis pela curadoria dos dados	Gestão dos dados de pesquisa	Infraestrutura
Gardner (2004)	Responsabilidade compartilhada	Gestão dos dados de pesquisa	Infraestrutura
Knight (2015)	Práticas de arquivamento (desconhecimento)	Gestão dos dados de pesquisa	Infraestrutura
Knight (2015)	Plano de gestão dos dados	Gestão dos dados de pesquisa	Infraestrutura
Thompson, Robertson e Greenberg (2015)	Preservação dos dados	Gestão dos dados de pesquisa	Infraestrutura
Thompson, Robertson e Greenberg (2015)	Preservação (padrões e políticas)	Gestão dos dados de pesquisa	Infraestrutura
Dallmeier-Tiessen et al. (2014)	Plano de preservação	Gestão dos dados de pesquisa	Infraestrutura
Dallmeier-Tiessen et al. (2014)	Preservação dos dados	Gestão dos dados de pesquisa	Infraestrutura
Dallmeier-Tiessen et al. (2014)	Financiamento para preservação	Gestão dos dados de pesquisa	Infraestrutura
Dallmeier-Tiessen et al. (2014)	Custos da preservação	Gestão dos dados de pesquisa	Infraestrutura
Dallmeier-Tiessen et al. (2014)	Qualidade dos dados	Gestão dos dados de pesquisa	Infraestrutura
Sturges et al. (2015)	Qualidade da comunicação científica	Gestão dos dados de pesquisa	Infraestrutura
Childs e Mcleod (2004)	Serviços de orientação aos pesquisadores	Gestão dos dados de pesquisa	Infraestrutura
Childs e Mcleod (2004)	Treinamento de profissionais	Gestão dos dados de pesquisa	Infraestrutura
Dallmeier-Tiessen et al. (2014)	Atuação de profissionais da informação	Gestão dos dados de pesquisa	Infraestrutura
Dallmeier-Tiessen et al. (2014)	Serviços de suporte	Gestão dos dados de pesquisa	Infraestrutura
Douglass et al. (2014)	Serviço de suporte aos pesquisadores	Gestão dos dados de pesquisa	Infraestrutura
Knight (2015)	Serviço para gestão dos dados	Gestão dos dados de pesquisa	Infraestrutura

Apêndices

Childs e Mcleod (2004)	Padrão de processos e procedimentos	Padrão	Infraestrutura
Childs e Mcleod (2004)	Indicações de padrões e políticas	Padrão	Infraestrutura
Whitmire, Boock e Sutton (2015)	Padrões de dados	Padrão	Infraestrutura
Thompson, Robertson e Greenberg (2015)	Padrões de metadados	Padrão	Infraestrutura
Dallmeier-Tiessen et al. (2014)	Interoperabilidade dos dados	Padrão	Infraestrutura
Dallmeier-Tiessen et al. (2014)	Padrões para a descrição dos dados	Padrão	Infraestrutura
Douglass et al. (2014)	Padrões	Padrão	Infraestrutura
Douglass et al. (2014)	Metadados (padrões)	Padrão	Infraestrutura
Douglass et al. (2014)	Interoperabilidade	Padrão	Infraestrutura
Knight (2015)	Padrão	Padrão	Infraestrutura
Knight (2015)	Interoperabilidade	Padrão	Infraestrutura
Knight (2015)	Formato dos dados	Padrão	Infraestrutura
Melero e Hernández-San-Miguel (2014)	Padrões	Padrão	Infraestrutura
Gardner (2004)	Tempo para tornar os dados disponíveis	Política de informação	Infraestrutura
Gardner (2004)	Divulgação exaustiva dos dados	Política de informação	Infraestrutura
Melero e Hernández-San-Miguel (2014)	Políticas de acesso aberto	Política de informação	Infraestrutura
Cox, Verbaan e Sen (2012)	Políticas de informação	Políticas de informação	Infraestrutura
Dallmeier-Tiessen et al. (2014)	Políticas mandatórias	Políticas de informação	Infraestrutura
Sturges et al. (2015)	Políticas editoriais	Políticas de informação	Infraestrutura
Losoff (2009)	Repositório de dados de pesquisa	Sistema da informação	Infraestrutura
Losoff (2009)	Periódicos para dados de pesquisa	Sistema da informação	Infraestrutura
Cox, Verbaan e Sen (2012)	Sistemas para a comunicação dos dados de pesquisa	Sistema da informação	Infraestrutura
Childs e Mcleod (2004)	Agrupamento dos dados em datasets	Sistema da informação	Infraestrutura
Whitmire, Boock e Sutton (2015)	Armazenamento dos dados (volume e preservação)	Sistema da informação	Infraestrutura
Whitmire, Boock e Sutton (2015)	Central de armazenamento dos dados	Sistema da informação	Infraestrutura
Thompson, Robertson e Greenberg (2015)	Acesso (mídias e repositórios)	Sistema da informação	Infraestrutura
Dallmeier-Tiessen et al. (2014)	Sistemas de descoberta dos dados	Sistema da informação	Infraestrutura
Dallmeier-Tiessen et al. (2014)	Falta de segurança	Sistema da informação	Infraestrutura
Dallmeier-Tiessen et al. (2014)	Integração dos sistemas	Sistema da informação	Infraestrutura
Dallmeier-Tiessen et al. (2014)	Relação dos dados com as publicações científicas	Sistema da informação	Infraestrutura
Douglass et al. (2014)	Motores de busca	Sistema da informação	Infraestrutura
Douglass et al. (2014)	Local de depósito	Sistema da informação	Infraestrutura
Knight (2015)	Segurança	Sistema da informação	Infraestrutura
Knight (2015)	Local de armazenamento	Sistema da informação	Infraestrutura

COSTA, M. **Fatores que influenciam a comunicação de dados de pesquisa sobre o vírus da zika, na perspectiva de pesquisadores.** 2017. Tese (Doutorado em Ciência da Informação) - Universidade de Brasília, Brasília, 2017.

Apêndices

Knight (2015)	Sistemas utilizados	Sistema da informação	Infraestrutura
Knight (2015)	Acesso (velocidade)	Sistema da informação	Infraestrutura
Sturges et al. (2015)	Repositório de dados de pesquisa	Sistemas de informação	Infraestrutura
Childs e Mcleod (2004)	Sistemas tecnológicos	Tecnologia	Infraestrutura
Childs e Mcleod (2004)	Infraestrutura tecnológica	Tecnologia	Infraestrutura

Apêndice B: Roteiro das entrevistas

Os pesquisadores entrevistados não serão identificados na entrevista. O acordo de confidencialidade dos dados foi firmado no início da entrevista com o objetivo de preservar a privacidade dos sujeitos e garantir a livre exposição das ideias.

Não é permitido o uso dos resultados da entrevista para outras finalidades que extrapolem essa pesquisa.

PERGUNTA DE ABERTURA

1. No conceito de **ciência aberta** há o pressuposto de que a pesquisa seja conduzida de um modo que outros pesquisadores possam colaborar e contribuir, em que os dados de pesquisa, as notas de laboratório e outros processos científicos estejam livremente disponíveis, com termos que permitam reuso, redistribuição e reprodução da pesquisa. Tendo como pano de fundo sua experiência como pesquisador, **qual sua percepção acerca do compartilhamento e reutilização de dados de pesquisa sob a ótica da ciência aberta?**

Gostaria que respondesse essa pergunta considerando **duas situações**:

- a) **você como produtor** de dados de pesquisa e agente promotor do compartilhamento para reutilização por outros pesquisadores;
- b) **você como pesquisador** que reutilizaria dados gerados por outros pesquisadores/pesquisas.

PRODUÇÃO

2. Nos projetos de pesquisa que você conduz, **como se dá**, em linhas gerais, a **produção/geração dos dados** que permitirão as descobertas científicas? Quais as **características** desses dados (tipos, volume, formas de análise, tecnologias envolvidas, formas de organização e armazenamento, forma de reutilização, etc)?
3. Considerando o tópico em que você atua e o sistema científico como um todo, quais aspectos você percebe como **influenciadores (estimuladores e/ou inibidores) de suas práticas de produção** de dados de pesquisa?
4. Considerando o contexto específico do **Vírus da zika**, há alguma **diferença ou particularidade** em relação à qualquer questão relacionada com a produção de dados? O fato do Vírus da zika ter sido classificado pela OMS como uma Emergência de Saúde Pública de Importância Internacional, influencia de algum modo sua percepção sobre as práticas de produção de dados em suas pesquisas ou de seus colegas?
5. Há qualquer aspecto contextual que **diferencia** a produção de dados sobre o Vírus da zika da produção de dados de pesquisa sobre **outros vírus ou doenças**?

COMPARTILHAMENTO

6. No que se refere ao avanço da ciência, o compartilhamento de dados de pesquisa tem sido considerado um pré-requisito em muitos países e áreas do conhecimento. Como você percebe essa questão? Você já teve alguma **experiência de compartilhamento** de dados de pesquisa **ou conhece** alguma iniciativa?
7. Considerando o tópico em que você atua e o sistema científico como um todo, é possível apontar aspectos percebidos como **estimuladores das práticas de compartilhamento** de dados de pesquisa? Quais?
8. Considerando o tópico em que você atua e o sistema científico como um todo, é possível apontar aspectos percebidos como **inibidores das práticas de compartilhamento** de dados de pesquisa? Quais?
9. Como você reagiria caso a agência que financia seus projetos tornasse o **compartilhamento** dos dados de suas pesquisas uma **condição para o financiamento**?
10. Como você reagiria caso sua **instituição** tornasse **compulsório o compartilhamento** dos dados de suas pesquisas?
11. Como você reagiria caso **editor** de um determinado periódico científico **exigisse** que o artigo submetido ou aceito para publicação fosse acompanhado dos **dados** da pesquisa que a gerou?
12. Considerando o contexto específico do **Vírus da zika**, há alguma **diferença ou particularidade** em relação à qualquer questão relacionada com o compartilhamento de dados da pesquisa que você realiza? O fato do Vírus da zika ter sido classificado pela OMS como uma Emergência de Saúde Pública de Importância Internacional, influencia de algum modo a sua percepção ou de seus colegas sobre o compartilhamento de dados de pesquisa?
13. Você considera que **os dados que você produz ou produzirá** sobre o Vírus da zika são **úteis** para a realização de **outras pesquisas**? **Estaria disposto a compartilhá-los? Por que?**
14. Há qualquer aspecto contextual que diferencia sua percepção sobre o compartilhamento de dados sobre o Vírus da zika do compartilhamento de dados de pesquisas sobre outros vírus ou doenças?

REUTILIZAÇÃO

15. É crescente a percepção o **avanço da ciência** pode ser **acelerado e otimizado** caso pesquisadores **reutilizassem dados** de pesquisa gerados por outros projetos. Como você percebe essa questão?
16. Você já teve alguma **experiência de reutilização** de dados que foram gerados por pesquisas conduzidas por outros pesquisadores? Se sim, quais fatores você pode apontar como **estimuladores dessa prática**? Se não, quais fatores você pode apontar como **inibidores dessa prática**?
17. Considerando o contexto específico do **Vírus da zika**, há alguma **diferença ou particularidade** em relação à qualquer questão relacionada com a reutilização

de dados gerados por outras pesquisas na condução da pesquisa que você realiza? O fato do Vírus da zika ter sido classificado pela OMS como uma Emergência de Saúde Pública de Importância Internacional, influencia de algum modo a sua percepção ou de seus colegas sobre a reutilização de dados de pesquisa?

18. Você considera que **os dados sobre o Vírus da zika produzidos por outros pesquisadores são úteis para a realização de suas pesquisas? Estaria disposto reutilizá-los?** Por que?
19. Há qualquer aspecto contextual que diferencia sua percepção sobre a reutilização de dados sobre o Vírus da zika da reutilização de dados de pesquisas sobre outros vírus ou doenças?

PERGUNTA DE FECHAMENTO

20. Além da OMS ter classificado o Vírus da zika como Emergência de Saúde Pública de Importância Internacional, a instituição reuniu lideranças representativas da comunidade científica, periódicos científicos, indústria, agências de fomento, governos de inúmeros países em Genebra em 2015 para identificar os principais obstáculos para o compartilhamento de dados e resultados de pesquisa em condições de emergências de saúde pública. É possível apontar **diferenças na atenção dada ao Vírus da zika e à outras doenças, sobretudo aquelas consideradas como negligenciadas?** Se sim, a **quais razões** você atribuiria essa diferença?