



UnB

Universidade de Brasília
Faculdade de Ciência da Informação
Programa de Pós-Graduação em Ciência da Informação

INGRID TORRES SCHIESSL

A relação entre os conjuntos de dados de pesquisa depositados em repositório de dados e os periódicos científicos

Brasília
2020



UnB

Universidade de Brasília
Faculdade de Ciência da Informação
Programa de Pós-Graduação em Ciência da Informação

INGRID TORRES SCHIESSL

**A relação entre os conjuntos de dados de pesquisa depositados em
repositório de dados e os periódicos científicos**

Dissertação de mestrado apresentada à banca examinadora do Programa de Pós-Graduação em Ciência da Informação da Faculdade de Ciência da Informação da Universidade de Brasília.

Orientação: Prof^a Dra. Fernanda Moreno.

Brasília
2020

Ficha catalográfica elaborada por Ingrid Torres Schiessl CRB1/3084

SCH332r Schiessl, Ingrid Torres

A relação entre os conjuntos de dados de pesquisa depositados em repositório de dados e os periódicos científicos / Ingrid Torres Schiessl; orientadora Fernanda Passini Moreno. -- Brasília, 2020.

132p.

Dissertação (Mestrado - Mestrado em Ciência da Informação) -- Universidade de Brasília, 2020.

1. Dados de pesquisa abertos. 2. Repositório de dados. 3. Ciência aberta. 4. Identificadores persistentes. 5. Gestão de dados de pesquisa. I. Moreno, Fernanda Passini, orient. II. Título.

FOLHA DE APROVAÇÃO

Título: “A RELAÇÃO ENTRE OS CONJUNTOS DE DADOS DE PESQUISA DEPOSITADOS EM REPOSITÓRIOS DE DADOS E OS PERIÓDICOS CIENTÍFICOS ”

Autor (a): Ingrid Torres Schiessl

Área de concentração: Gestão da Informação

Linha de pesquisa: Organização da Informação

Dissertação 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 **MESTRE** em Ciência da Informação.

Dissertação aprovada em: 04 de setembro 2020.

Presidente (UnB/PPGCINF): Fernanda Passini Moreno

Membro Externo (Unb/FCI): Michelli Pereira da Costa

Membro Interno (UnB/PPGCINF): João de Melo Maricato

Suplente (UnB/PPGCINF): Fernando César Lima Leite

Em 03/02/2020.



Documento assinado eletronicamente por **Fernanda Passini Moreno, Professor(a) de Magistério Superior da Faculdade de Ciência da Informação**, em 09/09/2020, às 08:29, conforme horário oficial de Brasília, com fundamento na Instrução da Reitoria 0003/2016 da Universidade de Brasília.



Documento assinado eletronicamente por **Joao de Melo Maricato, Professor(a) de Magistério Superior da Faculdade de Ciência da Informação**, em 16/09/2020, às 08:47, conforme horário oficial de Brasília, com fundamento na Instrução da Reitoria 0003/2016 da Universidade de Brasília.



Documento assinado eletronicamente por **Michelli Pereira da Costa, Professor(a) de Magistério Superior da Faculdade de Ciência da Informação**, em 16/09/2020, às 20:08, conforme horário oficial de Brasília, com fundamento na Instrução da Reitoria 0003/2016 da Universidade de Brasília.



A autenticidade deste documento pode ser conferida no site http://sei.unb.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **4949130** e o código CRC **2771B5D0**.

AGRADECIMENTOS

Agradeço à Prof^ª Fernanda Moreno pela orientação no desenvolvimento deste estudo. Suas colaborações, ao longo de três anos, foram essenciais para a minha formação acadêmica, profissional e crescimento pessoal. Agradeço à Prof^ª Michelli Costa e ao Prof^º João de Melo Maricato, por terem dedicado tempo à leitura do meu trabalho, contribuindo com minha pesquisa nas bancas de qualificação e defesa.

Agradeço também à Universidade de Brasília (UnB) por todos os recursos disponibilizados para a realização deste estudo e da minha formação desde a graduação. À secretaria do PPGCINF, em especial a Vívian (sempre tão solícita). Aos colegas de pós-graduação do PPGCINF, em especial a Raíssa da Veiga de Mênese e a Isabela Cardoso, que foram colegas de orientação e colaboraram para o desenvolvimento deste estudo.

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. Em especial, agradeço ao Milton Shintaku, que me incentivou antes e durante o mestrado, de modo que suas contribuições foram primordiais para esta pesquisa. Ao meu colega Ronnie Fagundes Brito, que me auxiliou no desenvolvimento do *script* para a coleta dos dados. Sem ele não seria possível sequer iniciar este estudo. À minha colega Rebeca Moura, que durante a apresentação do relatório intermediário se dispôs a ler todo o trabalho. As minhas colegas Cila, Jaque e Jan, que foram ouvintes nos momentos de fragilidade, mas também ofereceram palavras de conforto. À minha dupla, Ítalo, que foi se aventurar no mundo acadêmico germânico, mas enquanto perto foi amigo de trabalho e de vida acadêmica. Ao Rafael, que fez uma brilhante revisão linguística neste trabalho. Aos meus colegas Mari, Diego, Lucão, Luquinhas, Rafa, Nui, Guilherme e Mirele, que tornam a equipe da COTEC um lugar de enriquecimento pessoal e profissional.

Com amor, agradeço à minha família pelo suporte e carinho que recebi ao longo da vida. Em especial, à minha mãe e ao meu pai, que são as razões do que sou e do que já realizei. Minha irmã, companheira de profissão e melhor amiga. Ao meu companheiro Henrique, que permaneceu ao meu lado com carinho. Por fim, agradeço a todas as amigas, que me ouviram por horas falar sobre este trabalho, e ofereceram palavras de conforto e de encorajamento.

Estou entre aqueles que acham que a ciência tem uma grande beleza.

Marie Curie

RESUMO

Os dados ocupam um papel fundamental na Sociedade da Informação. Conforme a ciência se torna cada vez mais colaborativa, o compartilhamento de dados científicos torna-se mais importante. Tal compartilhamento não inclui apenas o depósito e a preservação, mas, principalmente, o acesso para uso e reúso de dados gerados no ciclo de vida da pesquisa. Nesse sentido, o estudo teve como objetivo investigar a relação, durante um período de dez anos, entre os conjuntos de dados depositados em repositórios de dados e os respectivos artigos científicos. Para isso, à luz da comunicação científica, foram estudados os conceitos de dados de pesquisa aberto, compartilhamento de dados, princípios e diretrizes para a gestão dos dados de pesquisa, repositório de dados abertos, periódicos científicos. A pesquisa se caracteriza como descritiva e apresenta abordagem quanti-qualitativa, de natureza básica, com método indutivo e com o horizonte temporal transversal, pois realizou a análise de conjuntos de dados em um determinado período. Para a coleta de dados utilizou-se um *script*, por meio de *Application Programming Interface* (API). Já a base selecionada para consulta aos repositórios de dados foi o DataCite, uma organização sem fins lucrativos que fornece gratuitamente identificadores persistentes para dados e resultados de pesquisa. Para a análise, obteve-se uma amostra de 67 de registros de conjuntos de dados depositados em repositórios de dados. A partir da análise, observou-se que os periódicos científicos das áreas entendidas como ciências rígidas – englobando, por exemplo, física e química – compartilham dados de pesquisa mais do que as outras áreas do conhecimento. Outro aspecto investigado foi a presença de políticas de compartilhamento de dados nas revistas científicas. Assim, constatou-se que apenas 33% não possuem políticas editoriais de comunicação de dados de pesquisa. Entre os periódicos científicos que apresentam a política de compartilhamento de dados de pesquisa, 58% obrigam o depósito em repositório de dados. Por fim, a investigação acerca da relação entre os dados depositados em repositórios e os periódicos científicos entende que, mesmo quando não há obrigatoriedade de compartilhamento, a presença de políticas de compartilhamento de dados nas revistas consiste em estímulo para o seu compartilhamento dos dados de pesquisa, uma vez que oferta diretrizes destinadas aos autores/pesquisadores sobre quais são as boas práticas para comunicação dos dados de pesquisa.

Palavra-chave: Dados de pesquisa abertos. Repositório de dados. Ciência aberta. Identificadores persistentes. Gestão de dados de pesquisa.

ABSTRACT

Data plays a fundamental role in the Information Society. As science becomes increasingly collaborative, sharing scientific data becomes more important. This sharing does not only include the deposit and preservation, but mainly the access for use and reuse of data generated in the research life cycle. The study aimed to investigate the relationship, during the ten-year period, between the data sets deposited in data repositories and the respective scientific articles. For this, in the light of scientific communication, the concepts of open research data, data sharing, principles and guidelines for the management of research data, open data repository, scientific journals were studied. The research is characterized as descriptive and presents a quantitative approach, of a basic nature, with an inductive method and with the transversal time horizon, as it performed the analysis of data sets in a given period. For data collection, a script was used, through the Application Programming Interface (API). The database selected for consulting the data repositories was DataCite, a non-profit organization that provides free persistent identifiers for data and search results. For the analysis, a sample of 67 records of data sets deposited in data repositories was obtained. From the analysis, it was observed that scientific journals in the areas understood as rigid sciences - encompassing, for example, physics and chemistry - share research data more than other areas of knowledge. Another aspect investigated was the presence of data sharing policies in scientific journals. Thus, it was found that only 33% do not have editorial policies for communicating research data. Among scientific journals that present the research data sharing policy, 58% require depositing in a data repository. Finally, research on the relationship between data deposited in repositories and scientific journals understands that, even when there is no mandatory sharing, the presence of data sharing policies in journals is an incentive for sharing research data, since it offers guidelines for authors / researchers on what are the best practices for communicating research data.

Keywords: Scientific open data. Data repository. Open science. Persistent identifiers. Research data management.

LISTA DE ILUSTRAÇÕES

FIGURAS

<i>Figura 1 - Modelo do processo de comunicação científica proposto por Garvey e Griffith</i>	22
<i>Figura 2 - Modelo de Hurd para o ano de 2020</i>	23
<i>Figura 3 - Modelo de comunicação científica para as comunidades de Ciências Sociais e Humanas</i>	24
<i>Figura 4 - Fluxo da informação científica e técnica do modelo UNISIST</i>	26
<i>Figura 5 - Diagrama de comunicar resultados</i>	27
<i>Figura 6 - Taxonomia da Ciência Aberta</i>	34
<i>Figura 7 - Classificação dos dados de pesquisa</i>	39
<i>Figura 8 - Licenças CC em ordem de mais a menos abertas</i>	51
<i>Figura 9 - Estrutura organizacional internacional de agências de publicação e agentes de publicação que finalmente levaram ao consórcio DataCite</i>	73
<i>Figura 10 - Organização da classificação de área do conhecimento do DOAJ</i>	85
<i>Figura 11 - Classe geral de assunto da ISCED</i>	89

QUADROS

<i>Quadro 1 - Categorias para se pensar a Ciência Aberta</i>	35
<i>Quadro 2 - Itens para ser considerados Open, segundo a OKF</i>	36
<i>Quadro 3 - Tipo de dados em relação à origem</i>	41
<i>Quadro 4 - Síntese dos tipos de dados de pesquisa</i>	42
<i>Quadro 5 - As tipologias dos metadados e suas definições</i>	43
<i>Quadro 6 - Guia dos princípios FAIR</i>	46
<i>Quadro 7 - Os princípios FAIR</i>	49
<i>Quadro 8 - Vantagens do compartilhamento dos dados abertos para a ciência e para a sociedade</i>	55
<i>Quadro 9 - Reusabilidade dos dados de pesquisa</i>	59
<i>Quadro 10 - Princípios propostos pela OECD para abertura dos dados de pesquisa financiadas</i>	62
<i>Quadro 11 - Conjunto dos princípios e recomendações para a comunicação dos dados de pesquisa</i>	66
<i>Quadro 12 - Os oito fatores que influenciam o compartilhamento de dados de pesquisa em diferentes áreas do conhecimento</i>	67
<i>Quadro 13 - Organizações com papel chave nos processos de incentivo aos repositórios de dados</i>	70
<i>Quadro 14 - Procedimentos metodológicos adotados no estudo conforme objetivos específicos</i>	81
<i>Quadro 15 - Metadados que indicam relacionamento e são parte do "superconjunto"</i>	84
<i>Quadro 16 - Correspondência entre a classificação do DOAJ e a Nomenclatura da Unesco para a área de Ciência Tecnologia (ISCED)</i>	89
<i>Quadro 17 - Lista de verificação de metadados</i>	93

Quadro 18 - Ficha de avaliação das revistas..... 94

Quadro 19 - Quantidade de compartilhamento de dados por periódico..... 104

GRÁFICOS

Gráfico 1 - Quantidade de compartilhamento por área do conhecimento 96

Gráfico 2 - Compartilhamento de dados por ano e área do conhecimento 98

Gráfico 3 - Presença do campo de licença de uso nos registros por área do conhecimento 100

Gráfico 4 - Avaliação dos registros de dados por área do conhecimento 102

Gráfico 5 - Presença de política de compartilhamento de dados por áreas do conhecimento 106

TABELAS

Tabela 1 - Quantidade de caracteres no campo dates por registro 82

Tabela 2 - Lista de registros coletados do DataCite por ano 86

Tabela 3 - Lista do tipo de material coletado no DataCite por registros 87

LISTA DE SIGLAS

AND	Australian National Data Service
API	Application Programming Interface
ARDC	Australian Research Data Commons
Bireme	Centro Latino-Americano e do Caribe de Informação em Ciências da Saúde
CI	Ciência da Informação
DCC	Digital Curation Centre
DOAJ	Directory of Open Access Journals
DOI	Digital Object Identifier
ERIC	Educational Resources Information Center
EU	European Union
FAIR	Findable, Accessible, Interoperable, and Reusable
Fapesp	Fundação de Amparo à Pesquisa do Estado de São Paulo
FOSTER	Facilitate Open Science Training For European Research
ICMJE	International Committee of Medical Journal Editors
IDF	International DOI Foundation
INSDC	International Nucleotide Sequence Database Collaboration
ISCED	Classificação Internacional de Educação
IURC	International Union of Crystallography
LOC	Library of Congress
NDLTD	Networked Digital Library of Theses and Dissertations
NERC	Natural Environment Research Council
NIH	National Institutes of Health
NSF	National Science Foundation
OECD	Organization for Economic Cooperation and Development
OKF	Open Knowledge Foundation
OMS	Organização Mundial da Saúde
OST	Office of Science and Technology
PLoS	The Public Library of Science
RCAAP	Projeto Repositório Científico de Acesso Aberto de Portugal
RIN	Research Information Network
RUCK	Research Councils United Kingdom
SciELO	Scientific Electronic Library On-line
SLAC	Stanford Linear Accelerator Center
TIB	German National Library of Science and Technology
TIC	Tecnologia de Informação e Comunicação
Unesco	Organização das Nações Unidas para a Educação, a Ciência e a Cultura
WGODS	Work Grouping Open Data in Science

SUMÁRIO

<u>1</u>	<u>INTRODUÇÃO</u>	<u>13</u>
1.1	Contextualização.....	13
1.2	Definição do problema	15
1.3	Objetivos	18
	1.3.1Objetivo geral.....	18
	1.3.2Objetivos específicos.....	18
1.4	Justificativa	18
<u>2</u>	<u>FUNDAMENTAÇÃO TEÓRICA.....</u>	<u>21</u>
2.1	Comunicação científica.....	21
	2.1.1Modelos de comunicação científica.....	22
	2.1.2Periódicos científicos.....	28
	2.1.3Ciência Aberta.....	33
2.2	Comunicação dos dados de pesquisa	38
	2.2.1Dados de pesquisa	38
	2.2.2Metadados.....	42
	2.2.3Os princípios FAIR (do inglês FAIR – Findable, Accessible, Interoperable, and Reusable) ..	45
	2.2.4Compartilhamento de dados de pesquisa	52
	2.2.5Repositório de dados.....	68
2.3	Sobre a fundamentação teórica.....	75
<u>3</u>	<u>METODOLOGIA.....</u>	<u>79</u>
3.1	Caracterização da pesquisa	79
3.2	Procedimentos metodológicos	80
	3.2.1Elaboração do script	81
	3.2.2Seleção da amostra.....	86
	3.2.3Análise do compartilhamento de dados pelas diferentes áreas do conhecimento	92

1.1.2	<i>Análise do nível de descrição dos registros de conjunto de dados</i>	92
3.2.4	<i>Roteiro de avaliação das políticas de compartilhamento de dados</i>	94
4	<u>RESULTADO E DISCUSSÕES</u>	95
4.1	Compartilhamento de dados nas áreas do conhecimento	95
4.2	Análise do nível de descrição dos registros de conjunto de dados	99
4.3	Avaliar as políticas de depósito de dados de pesquisa de periódicos com o intuito de verificar a relação entre as políticas e as áreas do conhecimento	103
5	<u>CONSIDERAÇÕES FINAIS</u>	111
	Contribuição do estudo	113
	Sugestão de estudos	113
	<u>REFERÊNCIAS</u>	114

1 INTRODUÇÃO

1.1 Contextualização

Os métodos de comunicação e transmissão de informações entre os cientistas sofreram mudanças ao longo do tempo. Destaca-se, nesse sentido, o advento da imprensa, que possibilitou maior acesso a textos impressos, contribuindo para a difusão mais rápida das pesquisas. Inúmeros acontecimentos transformaram os processos da comunicação científica, mas todas as mudanças foram motivadas pela necessidade de se comunicar da forma mais eficiente possível (MEADOWS, 1999).

Devido à utilização da web no cotidiano e, também, na comunicação científica, vivemos nas últimas décadas um período de transição para a segunda era da ciência. De acordo com Nielsen (2014), daqui a cem anos os historiadores olharão para trás e constatarão a existência de duas eras da ciência: a ciência antes da rede mundial de computadores conectados, também chamada de internet, e a ciência depois do surgimento dessa rede. Com a internet, tornou-se fácil e rápido o compartilhamento e envio de informações, bem como a colaboração entre cientistas.

Ao longo dos anos, os cientistas tiveram que se adaptar às novas maneiras de publicar e acessar os resultados, além de compartilhar informações. “O pesquisador envolvido com a comunicação científica, seja como autor, seja como editor, tem a responsabilidade de absorver tais inovações, isto é, inserir sua produção científica no novo modelo” (FERREIRA; TARGINO, 2010). Nesse sentido, destaca-se a Ciência Aberta, que retoma os princípios da ciência moderna e, entre outros objetivos, exige não apenas publicar, como também fazê-lo de forma aberta, sem cobrar nada aos usuários.

Segundo Albagli, Clinio e Raychtock (2014):

configura-se hoje um verdadeiro movimento de alcance internacional em favor da Ciência Aberta, a partir do suposto de que os modos atualmente dominantes de produção e de comunicação científica são inadequados, por estarem submetidos a mecanismos que criam obstáculos artificiais de várias ordens, especialmente legais e econômicos, à sua livre circulação e, logo, a seu avanço e difusão.

Na nova forma de comunicação científica, a palavra que norteia a Ciência Aberta é: colaboração. A publicação dos dados de pesquisa permite o reuso pelos pares; os códigos-abertos permitem o aprimoramento do software; a publicação dos cadernos de laboratório garante acesso aos erros e acertos em experimentos; os docentes podem criar materiais

educacionais de qualidade e gratuitos para discentes; o cidadão pode participar do processo científico. Importa destacar que não só o sistema tradicional de comunicação científica foi afetado, mas também as questões que lhe são inerentes e decorrentes, como propriedade intelectual, autoria coletiva, direito autoral, produtividade científica, citação etc. (PINHEIRO, 2014).

Esse momento estaria sob o domínio de uma “filosofia aberta”, definida por Costa e Moreira (2003) como “o uso de ferramentas, estratégias e metodologias que denotam um novo modelo de representar um igualmente novo processo de comunicação científica, ao mesmo tempo em que serve de base para interpretá-lo”.

O termo Ciência Aberta é comumente conhecido como guarda-chuva, que engloba diferentes significados, tipos de práticas e iniciativas, bem como envolve distintas perspectivas, pressupostos e implicações (ALBAGLI; CLINIO; RAYCHTOCK, 2014). Nesse sentido, Delfanti e Pitrelli (2015) incluem os seguintes movimentos: Acesso Aberto (*Open Access*); Dados abertos (*Open Data*); Ciência Cidadã (*Citizen Science*); Código Aberto (*Open Source*); Cadernos Abertos de Laboratório (*Open Notebooks*); Recursos Educacionais Abertos (*Open Educational Resources*); e Revisão por Pares Aberta (*Open Peer review*).

Pinheiro (2014) afirma que o regime de informação na comunicação científica foi alterado a partir das políticas de acesso livre à informação, o que impactou todos os atores participantes do processo. Machado (2015) relata que, em 2007, a definição de dados abertos em oito princípios veio dar mais força ao processo de abertura de dados e à ampliação dos usos e reúsos da informação, com impactos também na ciência.

O crescimento contínuo da quantidade de dados produzidos pelos diversos segmentos da sociedade – agências governamentais, instituições de pesquisa, indústria, entre outros – confere a esses recursos a condição de componente fundamental para a pesquisa científica moderna e os identifica, também, como parte dos fenômenos relacionados ao chamado Big Data, tão comentado ultimamente (SAYÃO; SALES, 2015).

O compartilhamento de dados é um recurso essencial para os princípios científicos de credibilidade, replicação e pesquisa adicional. Tal prática garante que métodos de pesquisa usados para produzir os resultados são conhecidos, e resultados incorretos podem ser retirados do corpo cumulativo de conhecimento (MUELLER-LANGER; ANDREOLI-VERSBACH, 2018). Ademais, estudos apontam que pesquisas que compartilham os dados

recebem mais citações do que aquelas que não o fazem (PIWOWAR; DAY; FRIDSMA, 2007; PIWOWAR; VISION, 2013).

Atualmente, as editoras científicas, entendendo a importância dos dados, exigem a disponibilização deles para avaliarem as publicações e reproduzirem os resultados. Como pioneiro, o *International Committee of Medical Journal Editors* (ICMJE) passou a solicitar disponibilização de ensaios sobre medicamentos. Outras editoras, como Nature e Science, publicaram seções sobre *Big Data* e propuseram políticas de depósitos de dados de pesquisa para os autores (KUM; AHALT; CARSEY, 2011). Existem periódicos que publicam artigos descrevendo os conjuntos de dados, as condições de produção e o interesse científico, que são chamados de *data papers* (AVENTURIER; ALENCAR, 2016).

A evolução e a disseminação do uso das Tecnologias de Informação e Comunicação (TICs) alteraram a forma como se dão as interações dos indivíduos com as informações científicas, o que, além de promover o desenvolvimento de ferramentas tecnológicas com uma alta capacidade de processamento e armazenamento, também facilitou o acesso e uso das publicações científicas. Esse cenário teve consequências imediatas e significativas para as pessoas e para as instituições. Porém, a quantidade de dados de pesquisa dispersos em diferentes contextos e domínios tornou-se um desafio para o gerenciamento das pesquisas e o reúso desses dados em outras pesquisas.

1.2 Definição do problema

À medida que o mercado da comunicação acadêmica continua evoluindo, alguns indicadores sugerem que a unidade de moeda da informação está mudando de um foco primário, nos artigos de periódicos, para uma ênfase mais ampla nos elementos-chave da comunicação acadêmica, isto é, os conjuntos de dados (DAVIS; VICKERY, 2007). Sendo assim, conforme afirmam Carvalho e Leite (2019), os dados de pesquisa têm introduzido mudanças em relação ao ciclo tradicional da comunicação científica, na medida em que possibilitam aos pesquisadores diferentes maneiras para desenvolver e disseminar suas pesquisas.

A preocupação com os dados de pesquisa levanta questões na comunicação científica, com vistas a sobrepor as antigas convenções do que são os produtos da investigação científica. A comunidade científica, inserida na chamada Ciência Aberta, vivencia uma nova demanda, que é a disponibilização aberta e inteligível dos dados de pesquisa (CARVALHO; LEITE, 2019).

Disponibilizar dados científicos tratados para pesquisadores não é uma questão nova na CI (SALES; SAYÃO, 2012). Os autores lembram que o Museu Paraense Emílio Goeldi, em fins da década de 1980 e início de 1990, desenvolveu o PRIMATAM, projeto ligado ao Núcleo de Primatologia cujo tratamento dos dados de pesquisa resultou em um catálogo impresso. Outro exemplo é o projeto Genoma, que abriu seus dados à comunicação científica e ao público. E, desde a década de 1980, o *International Nucleotide Sequence Database Collaboration* (INSDC) mantém o Genbank, o qual é um conjunto de bases de dados que trata de informações sobre sequenciamento genômico das mais diversas espécies. Importa ressaltar, conforme afirmam Ruttenberg e Risbeth (1994), que as iniciativas para a comunicação dos dados de pesquisa foram desenvolvidas ainda na década de 1950. Na década de 1970, por sua vez, o tema já é abordado na literatura da ciência da informação (SØNDERGAARD; ANDERSEN; HJØRLAND, 2003).

Foi na década de 1990, com a expansão mundial da Internet, que começaram a se proliferar os bancos de dados de artigos científicos de livre acesso. Como destaque, tem-se em 1996 o surgimento da *Networked Digital Library of Theses and Dissertations* (NDLTD),¹ que passou a ser o maior banco de teses e dissertações do mundo. Como iniciativa da América Latina há a base de periódicos *Scientific Electronic Library On-line* (SciELO), criada pelo Centro Latino-Americano e do Caribe de Informação em Ciências da Saúde (Bireme) com apoio da Fundação de Amparo à Pesquisa do Estado de São Paulo (Fapesp). Os repositórios científicos também desempenharam importante papel no acesso aberto, permitindo a disponibilização de artigos e documentação de pesquisa produzidos nas universidades e nos centros de investigação.

Ao longo do tempo, estudiosos indicam que o acesso ao conhecimento e à informação são essenciais para o desenvolvimento humano, social e econômico. No contexto político brasileiro há forte consenso no sentido de considerar o direito de acesso à informação como algo fundamental (MACHADO, 2015). Conseqüentemente, a abertura dos dados científicos representa enormes benefícios para a sociedade. Além disso, há aumento no número de iniciativas científicas *open*, as quais levam ao surgimento de novos paradigmas para produção e distribuição do conhecimento, o que de certa maneira destaca os dados de pesquisa na comunicação científica.

¹ Disponível em: <http://www.ndltd.org>.

O que facilmente se nota é o fato de que dados de pesquisa têm assumido um maior protagonismo na ciência, pois “passaram a ser valorizados como ativos de pesquisa autônomos, de alto valor intrínseco e publicáveis” (CURTY; AVENTURIER, 2017). O estabelecimento de métricas para avaliar a significância e o impacto da produção e publicação de dados torna-se cada vez imprescindível, uma vez que o impacto também vem sendo considerado como a moeda do mundo acadêmico (NASSI-CALÒ, 2015).

Para Machado (2015, p. 202), o debate sobre o acesso aberto diz respeito a dados abertos. O autor afirma que, atualmente, difundem-se protocolos, formatos e plataformas que permitem maior interoperabilidade, processamento, cruzamento e reuso da informação. Assim, conhecer as relações e processos acerca do compartilhamento de dados de pesquisa se faz importante, visto que novos modelos, processos e paradigmas são criados e uma nova forma de fazer ciência vem se consolidando no mundo digital.

No contexto do compartilhamento de dados percebe-se que a transformação ocorre para pesquisas, autores/pesquisadores e disciplinas e/ou área do conhecimento. Esse novo paradigma, vivido na ciência, impõe desafios e cria oportunidades. A pesquisa realizada por Stuart *et al.* (2018, p. 4) demonstra que os principais desafios enfrentados pelos pesquisadores para o compartilhamento de dados são: tratamento do dados de uma forma apresentável e útil; falta de informação sobre direitos autorais e licenciamento; não saber qual repositório usar; falta de tempo para depositar dados; e custos de compartilhamento de dados.

No que diz respeito às pesquisas, o ato de compartilhar dados permite a validação da metodologia e a elaboração de pesquisas relacionadas ao estudo. Dessa forma, maiores insumos sobre o tema podem ser produzidos e publicados. Além disso, o reuso de dados reduz o tempo de pesquisa, já que a coleta não precisa ser realizada.

Em relação às áreas do conhecimento, a fim de respeitar os critérios éticos, o compartilhamento de dados exige o comprometimento para a elaboração de políticas e normas de disponibilização, isto porque as especificidades de cada área impactam na forma de coleta e análise dos dados e, por conseguinte, no próprio dado. Ademais, a abertura dos dados é mais um aliado na detecção de plágios e fraudes, além de promover a interdisciplinaridade, pois a partir do cruzamento de dados é possível gerar novos estudos. Por fim, a abertura dos dados viabiliza a transparência da pesquisa, o que facilita o apoio de agências de fomento à pesquisa.

Nesse contexto, o problema de pesquisa pode ser sintetizado pela seguinte pergunta: **“Como se caracteriza o cenário sobre o compartilhamento de dados de pesquisa que tenha relação com artigos científicos publicados?”**.

1.3 Objetivos

1.3.1 Objetivo geral

Analisar a relação dos conjuntos de dados depositados em repositórios de dados com os periódicos científicos.

1.3.2 Objetivos específicos

- Compreender como as diferentes áreas do conhecimento compartilham dados;
- Analisar o nível de descrição dos registros de dados pelas diferentes áreas do conhecimento;
- Avaliar as políticas de depósito de dados de pesquisa de periódicos com intuito de verificar a relação entre as políticas e as áreas do conhecimento;

1.4 Justificativa

Compartilhar dados não é um tópico novo, tanto que na década de 1970 o modelo de comunicação científica UNISIST discutia os dados tabulares (UNESCO, 1971). Além disso, a pesquisa científica, no século XXI, se apresenta mais intensiva em dados e colaborativa do que no passado. Inclusive, como já mencionado, o projeto Genoma desde a década de 1980 estimula a disponibilização de conjuntos de bases de dados para os pesquisadores integrantes do consórcio.

Em 2003, a Organização das Nações Unidas para a Educação, a Ciência e a Cultura (*United Nations Educational, Scientific and Cultural Organization – Unesco*) publicou a carta para a preservação da herança digital. Neste documento, a instituição afirma que os dados digitais de pesquisa são considerados parte da herança cultural digital da humanidade e, assim sendo, as nações devem preservá-los com o intuito de garantir acesso futuro (UNESCO, 2003).

Novas competências associadas ao trabalho com os dados de pesquisas surgem para corroborar com o descobrimento a partir dos volumes de dados. As pesquisas sobre o tema se encontram dentro da Ciência da Informação (CI), pois na literatura se destaca a importância do profissional da informação, o qual impacta a Comunicação Científica, o

desenvolvimento de coleções, a criação e gestão de repositórios de dados e os padrões de metadados (ALLARD; MACK; FELTNER-REICHERT, 2005; CHOUDHURY, 2008; STEINHART, 2006).

Ao abordar questões de acesso aos dados de pesquisa, adentra-se no estudo da comunicação científica. Na CI, os estudos pioneiros sobre o tema são, em sua maioria, abordagens quantitativas que enfatizam questões acerca da produção e do uso da literatura científica. A criação das políticas de depósito de dados em periódicos científicos contribuirá para a literatura nacional, visto que a abordagem do tema no país é recente e corrobora a discussão sobre o compartilhamento e a curadoria de dados, que são as pesquisas expoentes na literatura nacional e internacional.

A falta de informações a respeito da valorização dos dados é um dos obstáculos que impede o avanço da disponibilização deles. Alguns dados são sensíveis, como, por exemplo, informações confidenciais dos pacientes ou estratégias que não devem ser publicadas imediatamente. Além disso, são escassos os profissionais especializados em curadoria e disponibilização de dados de pesquisas, uma vez que se trata de uma atividade recente. Nesse caso, os cientistas da informação e bibliotecários têm papel fundamental e reconhecido pelas diferentes organizações na área.

Em seu estudo sobre o compartilhamento de dados de pesquisa na área da genética, Piwowar (2011) afirma que, apesar de ser uma área de conhecimento com políticas, repositórios e padrões bem desenvolvidos, os níveis de compartilhamento de dados de pesquisa são baixos e aumentam lentamente. Logo, os dados estão menos disponíveis em áreas onde poderiam causar maior impacto, como nos casos de pesquisas sobre câncer em seres humanos. Em sua pesquisa, a autora discute a necessidade de estudos para compreender quais fatores estão associados ao compartilhamento de dados, visto que em primeiro momento identificou-se que aspectos como a disciplina da pesquisa, o financiamento, a publicação, a política institucional e a ideologia do autor podem ser associados à probabilidade de compartilhamento de dados.

De fato, em muitos casos os cientistas não dispõem de tempo hábil e/ou conhecimento para a gestão dos próprios dados, formas de valoração dos dados e interesse em compartilhá-los. Embora existam incentivos das instituições, agências de fomento e políticas dos periódicos em geral, os pesquisadores não disponibilizam os dados de pesquisa, mas admitem que compartilhá-los colabora para o progresso científico (ANDREOLI-VERSBACH; MUELLER-LANGER, 2014).

Pesquisadores coletam dados para diferentes propósitos utilizando vários métodos. Tanto os propósitos quanto os métodos influenciam na maneira como os cientistas consideram, o que é dado e sob quais condições eles irão compartilhar os dados. Assim, entender as práticas, os problemas e as políticas para dados é algo em expansão no campo da CI (BORGMAN, 2012).

Contudo, entender e analisar o cenário que se constrói em relação ao compartilhamento de dados é importante para ciência, pois é preciso conhecer as necessidades para dar início à implementação de estruturas adequadas para o compartilhamento de dados. Nas comunidades científicas, a discussão sobre a explosão na produção de dados cresce cada vez mais (RODRIGUES *et al.*, 2010). Os mesmos autores afirmam que haverá aumento do volume de dados devido ao crescimento das atividades de pesquisa ou em decorrência dos novos métodos e instrumentos de pesquisa e registro, os quais originam volumes de dados cada vez maiores.

Portanto, uma vez que o fenômeno do compartilhamento de dados de pesquisa tem sido identificado como elemento capaz de melhorar o fluxo de comunicação e otimizar os processos da ciência, caracterizar o cenário do compartilhamento de dados de pesquisas publicadas em periódicos científicos – isto por meio da análise da relação dos conjuntos de dados depositados em repositórios com os respectivos periódicos científicos – trará subsídios à comunidade brasileira nas discussões acerca da temática.

2 FUNDAMENTAÇÃO TEÓRICA

Este capítulo apresentará as definições e a discussão sobre os principais temas desta dissertação, além de sua articulação com a literatura escolhida como base para o desenvolvimento do trabalho.

2.1 Comunicação científica

Comunicar de modo eficiente os resultados de pesquisa aos pares e aos leigos incide diretamente no progresso da ciência (MUELLER, 2000). Nesse sentido, a CI tem estudado, desde a sua concepção, questões relacionadas com a comunicação científica, a qual surgiu em virtude da preocupação dos cientistas, dos tecnólogos e dos documentalistas com os fluxos da informação em ciência e tecnologia (COSTA; LEITE, 2018).

A comunicação científica pode ser definida, segundo Garvey (1979), como “todo o conjunto de atividades associadas à produção, disseminação e uso de informações, desde o momento em que o cientista concebe a ideia da pesquisa até que os resultados dessa pesquisa sejam aceitas como conhecimento científico”. De acordo com Roosendaal e Geurts (1997), a comunicação científica possui quatro funções principais de registro de autoria, certificação, circulação e arquivamento. Elas são alistadas por Costa (2014), respectivamente, como a função relacionada com a propriedade intelectual; a garantia da qualidade e validação dos conhecimentos gerados; a promoção da acessibilidade aos resultados de pesquisa; e a preservação para uso futuro. Nesse sentido, para a autora, 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” (COSTA, 2014).

Para Ziman (1984), a comunicação científica desempenha papel primordial no desenvolvimento do conhecimento científico, pois um conhecimento só se torna científico se for comunicado. Por sua vez, Meadows (1999) argumenta que o conhecimento científico é gerado a partir da relação entre pesquisa e comunicação. Desse modo, Shintaku e Costa (2018), levando em consideração essas afirmações, argumentam que a tríade conhecimento científico-pesquisa-comunicação científica é uma relação de interdependência funcional, isto é, uma não existe sem a outra, haja vista que o conhecimento resultante de pesquisa só ganha status de conhecimento científico por intermédio da comunicação. Sob essas circunstâncias, a comunicação perpassa todas as fases da pesquisa, inclusive a origem e o fim, porque a própria pesquisa se origina de uma comunicação e se encerra na disseminação dos resultados.

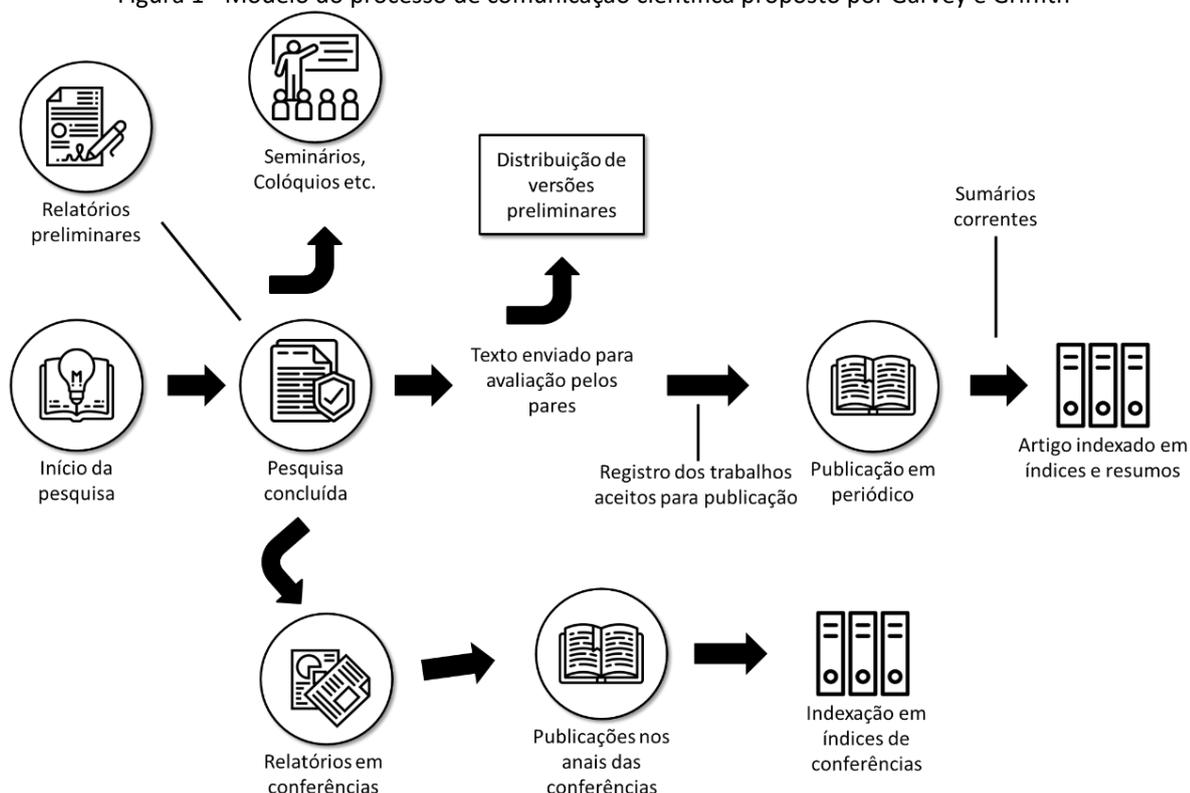
De acordo com Leite (2006), a comunicação científica “é um processo fundamental e inseparável da atividade científica”. Por isso, as diferentes áreas do conhecimento formalizam as práticas para circulação e trocas de informação entre os pesquisadores e o público geral (PINTO; COSTA, 2018). Como consequência, para compreender os processos e as práticas foram elaborados diversos modelos sobre a comunicação da ciência. Contudo, tais modelos podem variar conforme os interlocutores, as características e os meios que estão ao dispor de quem produz conhecimento (PINTO; COSTA, 2018).

2.1.1 Modelos de comunicação científica

As áreas do conhecimento concebem padrões e normas que, em alguns aspectos, são semelhantes e, em outros, diferentes – isto é, existem formas diferentes de “fazer ciência” e, provavelmente, de comunicar a pesquisa. Nesse sentido, em 1979, Kuhn (1970) afirma que no processo de pesquisa científica a comunidade não funciona de maneira homogênea e seu comportamento está relacionado com uma divisão em grupos.

O modelo de comunicação científica de Garvey e Griffith (Figura 1) é um dos pioneiros. Foi desenvolvido para uma área do conhecimento específica (a Psicologia), de modo que considerava os periódicos impressos (COSTA, 1999).

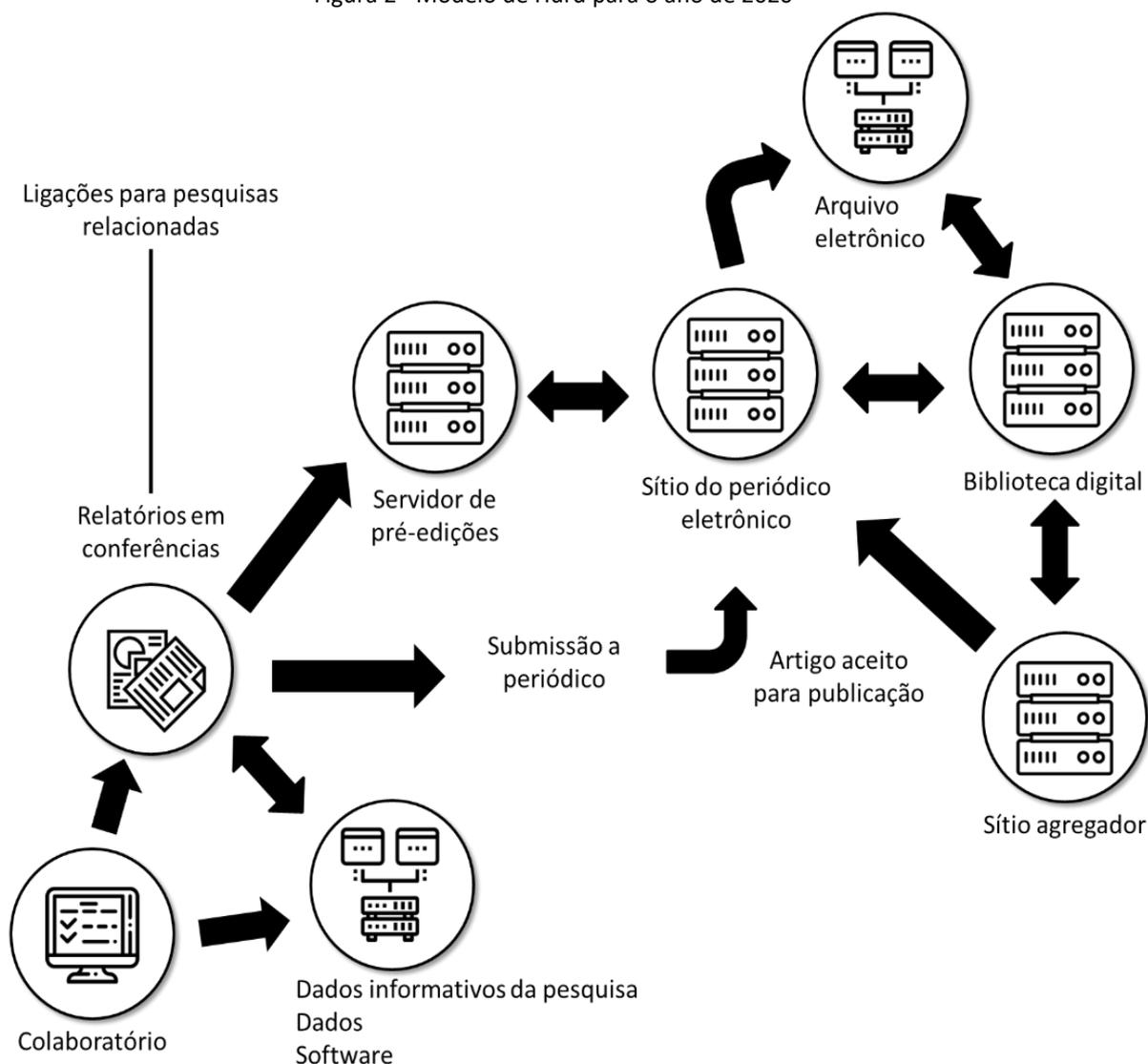
Figura 1 - Modelo do processo de comunicação científica proposto por Garvey e Griffith



Fonte: Pinto e Costa (2018, p. 146) adaptado de Hurd (2000).

A partir de meados da década de 1980, com os avanços tecnológicos e o surgimento dos periódicos eletrônicos, novos modelos foram apresentados. Hurd (2000) acrescentou ao modelo de Garvey e Griffith o componente eletrônico (Figura 2). Assim, os canais informais tradicionais, como o telefone e o contato pessoal são substituídos pelo correio eletrônico ou pelas listas de discussão (LEITE, 2006).

Figura 2 - Modelo de Hurd para o ano de 2020



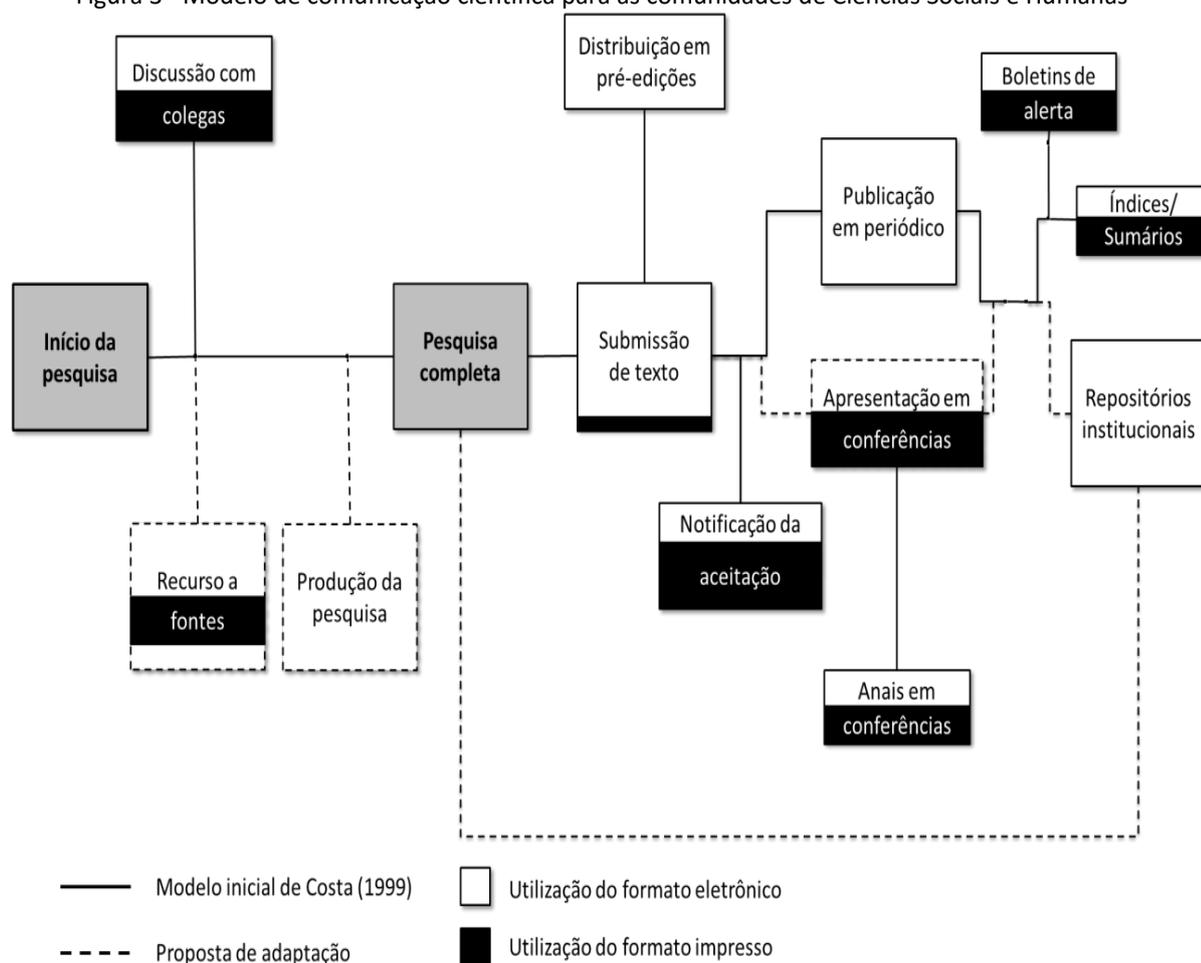
Fonte: Pinto e Costa (2018, p. 148) adaptado de Hurd (2000).

Segundo Leite (2006), tal modelo justifica-se pelo fato de os canais eletrônicos serem excelentes para contatos entre pesquisadores quando separados por grandes distâncias. Importa ressaltar que, em razão das características e dos modos de produzir e comunicar o saber, as diferentes disciplinas das ciências adotam formatos que melhor satisfazem suas necessidades (PINTO; COSTA, 2018).

Levando em consideração os modelos de Garvey e Griffith e Hurd, Costa (1999) propõe um modelo híbrido, pois segundo a autora era necessário fazer a “ponte” entre o impresso e o digital. O estudo desenvolvido pela autora ainda investiga comunicação na área das Ciências Sociais, e esclarece que, apesar da preponderância dos meios eletrônicos em determinadas fases, a comunicação formal ainda era realizada com base no modelo tradicional, em que livros e periódicos impressos se impõem aos novos formatos digitais (COSTA, 1999).

Os autores Pinto e Costa (2018) afirmam, em seu estudo, que o modelo de comunicação que mais se aproxima da realidade é o de Costa (1999). Logo, os autores propõem um novo modelo de comunicação para as comunidades de Ciências Sociais e Humanas, baseado no modelo da autora (Figura 3).

Figura 3 - Modelo de comunicação científica para as comunidades de Ciências Sociais e Humanas



Fonte: Pinto e Costa (2018, p. 155).

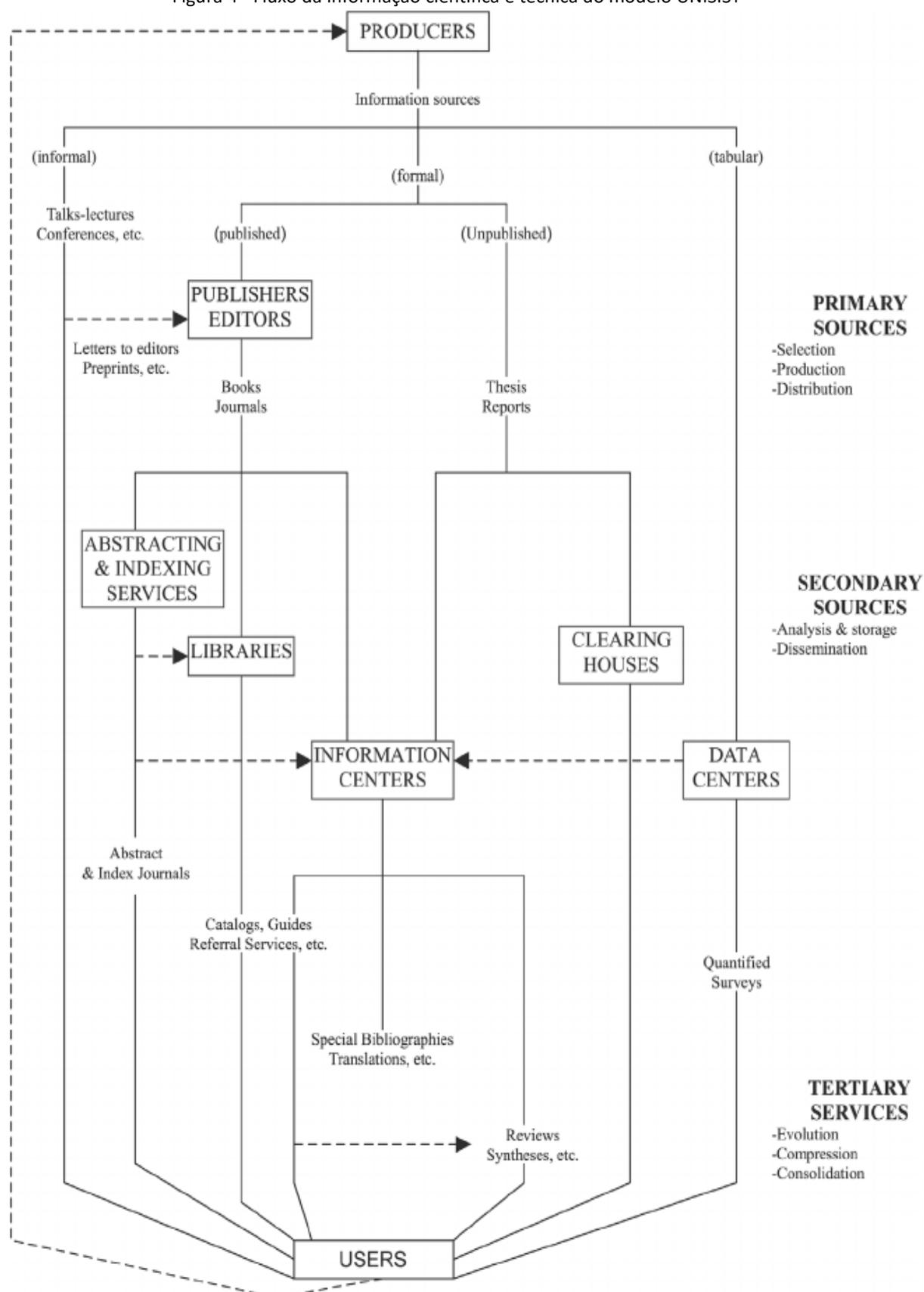
Outro modelo de comunicação científica é o UNISIST, criado em 1971, que oferece a ótica sóciotécnica sobre as atividades de comunicação científica (SØNDERGAARD;

ANDERSEN; HJØRLAND, 2003). Segundo os autores, o modelo define um sistema formado por várias unidades organizacionais e documentárias, em que todas colaboram para o compartilhamento de atividades na comunicação científica. Esse modelo dá enfoque para a comunicação da informação entre produtores e usuários de conhecimento. De fato, desde a concepção do modelo UNISIST, o fluxo da comunicação científica vem mudando devido ao uso e ao impacto dos canais de comunicação baseados na internet (SØNDERGAARD; ANDERSEN; HJØRLAND, 2003). Por esse motivo, a partir do ponto de vista do uso da internet, os autores atualizaram o modelo UNISIST. Assim, cria-se um modelo baseado na divisão nos tipos de canais, isto é, o canal formal, o informal e o tabular (LEITE, 2011). De acordo com Leite (2011):

os canais informais de comunicação “são utilizados quando produtores e usuários conhecem um ao outro e trocam informação por meio de correspondências pessoais, distribuição de pré-prints, indicações de leituras, etc.”. Por outro lado, os canais formais de comunicação são divididos em documentos publicados e documentos não publicados. Por fim, os canais tabulares comunicam dados técnicos e científicos. É por meio deles que os dados são apresentados em forma de tabelas, ao contrário do fluxo linear do texto falado ou escrito, como afirma o autor.

Ainda sobre o modelo UNISIST, cabe destacar a discussão sobre a comunicação de dados de pesquisa, representada pelos canais tabulares, isto é, unidades organizacionais denominadas de centros de dados seriam as responsáveis por essa comunicação dos dados (SØNDERGAARD; ANDERSEN; HJØRLAND, 2003). No contexto em que o modelo foi proposto, conforme Costa e Leite (2017, p. 91), 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.

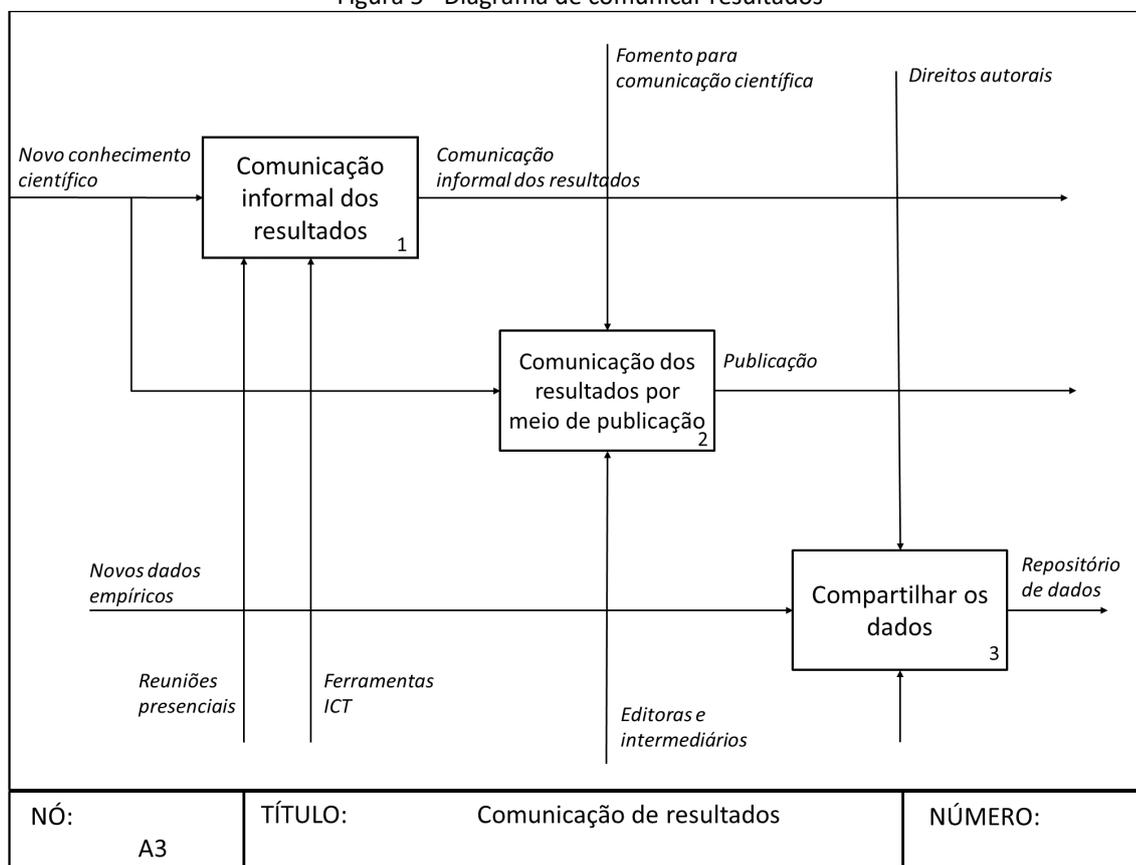
Figura 4 - Fluxo da informação científica e técnica do modelo UNISIST



Fonte: (SØNDERGAARD; ANDERSEN; HJØRLAND, 2003)

Dentre os modelos de comunicação, há o proposto por Björk (2007), que também divide o processo de comunicação científica em uma parte informal e outra formal, assim como o UNISIST. No modelo, é adotada uma visão mais funcional, na qual os estágios antes da publicação formal (documentos de trabalho, disponibilização de *preprints* etc.) são agrupados com a publicação formal tradicional. O diagrama elaborado pelo autor leva em consideração o fato de que os cientistas não apenas publicam documentos textuais, como também dados e modelos. Exemplos destes últimos podem incluir dados de observação astronômica, modelos de realidade virtual de artefatos históricos, gráficos de genoma e código de computador.

Figura 5 - Diagrama de comunicar resultados



Fonte: (BJÖRK, 2007).

Björk (2007) representa a publicação de dados e modelos no diagrama, pois entende que são partes essenciais do processo de gerenciamento do conhecimento científico. Contudo, o autor não se aprofunda nesses processos, afirmando apenas que, ao detalhá-los, o nível de complexidade do modelo aumentará. Para ele:

O escopo do modelo é toda a cadeia de valor da comunicação científica, desde a pesquisa inicial até a assimilação dos resultados da pesquisa para melhorar a vida cotidiana. O modelo trata da comunicação informal e formal, bem como da publicação de dados, mas o foco principal é modelar a publicação e a indexação de

artigos de periódicos tradicionais revisados por pares, bem como as atividades dos leitores para descobrir sobre eles e acessá-los. Os novos modelos de negócios e funções paralelas habilitadas pela Internet, como periódicos de acesso aberto e repositórios de impressão eletrônica, também estão em foco (BJÖRK, 2007).

Entre os modelos citados, é possível observar a presença do artigo de periódico científico como um canal de comunicação dos resultados de pesquisa, de tal maneira que na literatura sobre o tema discute-se a importância dos artigos de periódico como um dos principais meios para a disseminação de informações científicas (MEADOWS, 1999; MUELLER; PASSOS, 2000; MUELLER, 2000; PINTO; COSTA, 2018; TENOPIR; KING, 2001). Contudo, importa ressaltar que áreas do conhecimento divergem quanto à predileção pelos periódicos como meio de comunicação científica.

2.1.2 Periódicos científicos

De acordo com Mueller (2000), no século XVII houve uma mudança no meio científico, uma vez que a dedução deixou de ser aceita como método principal de pesquisa e a comunidade científica começou a exigir evidências baseadas na observação e na experiência empírica para que os conhecimentos resultantes pudessem ser considerados científicos. A partir desse momento ocorreu o nascimento da ciência moderna e, como consequência, alterou-se a forma de comunicação científica. Com o advento da ciência moderna, a necessidade de comunicar os resultados de pesquisa de forma rápida e precisa ganhou importância, isto porque a forma de divulgação formal das pesquisas costumava ser feita em livros e tratados (MUELLER, 2000). Daí surgiu a necessidade de criar um meio de comunicação: o periódico científico.

Os primeiros periódicos científicos de que se tem notícia são o *Journal de Sçavans*, fundado pelo francês Denis de Sallo, cujo primeiro fascículo foi publicado em 5 de janeiro de 1665; e o *Philosophical Transactions of the Royal Society*, que surgiu em Londres, dedicado exclusivamente ao registro das experiências científicas e publicado menos de três meses após o *Journal de Sçavans* (MUELLER, 2000, p. 74). A primeira publicação, o *Journal de Sçavans*, por muito tempo não aparecia regularmente, mesclando artigos noticiosos e científicos (MARTINS, 2003, p. 22). Já a publicação inglesa, por sua vez, dada a delicada situação político-religiosa por que passava o país, privilegiava relatos “experimentais” no lugar de reflexões com viés político ou religioso. Porém, como afirma Meadows (1999), além da temática experimental o *Philosophical Transactions* apresentou algumas características que os periódicos científicos praticam até hoje, tais como: tornar

público os relatos dos membros da sociedade, receber colaboração de não sócios, cobrar uma assinatura dos leitores e instituir regularidade aos fascículos publicados.

Os dois periódicos inauguraram uma forma de comunicação impressa mais rápida e barata, hoje denominada periódico científico (AULER, 2019). Antes, os veículos de comunicação científica por escrito eram livros, cartas e resumos das reuniões de uma sociedade ou academia. O novo tipo de publicação, chamado periódico, era diferente e, de forma mais rápida, reunia material diversificado (como relatos, resumos de livros, notícias), agrupava conteúdo suficiente para preencher um fascículo e era cofinanciado por assinaturas e pela sociedade ou academia responsável por ele (GUANAES; GUIMARÃES, 2012; MARTINS, 2003; STUMPF, 1996; ZIMAN, 1969).

Observa-se que os artigos de periódicos científicos, que apresentavam um discurso cuidadoso e impessoal, substituíram a informalidade das cartas e a oralidade das apresentações em reuniões acadêmicas (AULER, 2019). Os artigos necessitavam ser redigidos em formato que pudesse ser interpretado por todos, como afirmam Ziman (1969) e Stumpf (1996). Ziman (1969), porém, ressalta que as condições e os métodos dos experimentos precisavam estar descritos de modo a tornar possível a sua reprodução por outros pesquisadores.

O periódico científico se consolidou como um meio de comunicação científica. Por isso, tem como funções garantir a memória da ciência, apontar o grau de evolução da ciência, estabelecer a propriedade intelectual, legitimar novos campos de estudos e disciplinas, constituir-se como fonte para o início de novas pesquisas, entre outras (BIOJONE, 2003; MARCHIORI; ADAMI, 2005; MEADOWS, 1999; STUMPF, 1996).

Como a tecnologia modificou os meios de comunicação científica, conseqüentemente os periódicos também foram impactados. De acordo com Mueller (2000, p. 76-77), os problemas com o modelo tradicional de periódico científico são:

[...]• demora na publicação do artigo que, às vezes, chega a ser de um ano após o recebimento do original pelo editor; • custos altos de aquisição e manutenção de coleções atualizadas; • rigidez do formato impresso em papel, quando se compara com a versatilidade dos formatos eletrônicos; dificuldade, para o pesquisador, em saber o que de seu interesse está sendo publicado, pois são muitos os periódicos e pouco eficientes os instrumentos de identificação e busca; • dificuldade, para o pesquisador, em ter acesso a artigos que lhe interessam, pois mesmo sabendo que um novo artigo de seu interesse foi publicado, nem sempre sua biblioteca assina o periódico que o publicou ou consegue obter uma cópia desse artigo com a rapidez suficiente.

A causa desses problemas é discutida pela autora, que identifica três fatores ligados entre si: a proliferação de periódicos, a dispersão de artigos e o custo de atualização de coleções. O surgimento de novos periódicos traz consigo um aumento dos artigos, isto é, o mesmo assunto discorrido entre muitos títulos. Consequentemente, há a necessidade de aumentar as coleções de periódicos por meio de aquisição de novas obras, a fim de que se possa abarcar os estudos sobre cada tema de forma ampla, o que eleva em demasia o custo da atualização de coleções. Ainda para a autora, esses fatores foram reações ao crescimento natural da ciência, às regras da comunidade científica, que em certa medida impõem a publicação aos pesquisadores para progressão na carreira e financiamento da pesquisa, e aos cortes de orçamentos sofridos pelas bibliotecas em todo o mundo.

O desenvolvimento muito rápido da Internet e, em particular, dos serviços disponíveis na rede desde 1994, modificaram profundamente o acesso à informação, inclusive os periódicos. Esse meio de comunicação formal migrou para o suporte eletrônico, o que modificou tanto o mercado editorial da comunicação científica como as estratégias de visibilidade do conhecimento (WEITZEL, 2014). Em consequência, de acordo com Meadows (1999), os artigos de periódicos também se modificaram. Tais mudanças ocorreram porque era necessário melhorar a eficiência na comunicação entre pesquisadores de comunidades que cresceram e se tornaram mais complexas.

Nesse sentido, diante do cenário elucidado por Curty e Aventurier (2017), no qual os dados são considerados produtos de pesquisa autônomos com alto valor intrínseco, e as coleções de dados estão evoluindo e se popularizando entre os membros da comunidade científica mundo afora. Observa-se o surgimento dos periódicos de dados (*data journals*), apresentando os *data papers* como um novo tipo de publicação científica dedicados exclusivamente à descrição dos dados de forma estruturada e legível para humanos (PENEV *et al.*, 2013).

2.1.2.1 Os periódicos de dados (*data journals*)

Observa-se o surgimento de meios para comunicar os dados. Em 2014, o *Nature Publication Group* lança o primeiro periódico de acesso aberto a dados de pesquisa, o *Scientific Data*, cujo foco é a descrição de conjuntos de dados de pesquisa que têm como objetivo promover o compartilhamento e o reuso dos dados (CAVALCANTI; SALES, 2017).

Scientific Data é um novo periódico de acesso aberto, apenas online, para descrições de conjuntos de dados científicos valiosos. Nossos artigos, conhecidos como *Data Descriptors*, combinam conteúdo narrativo tradicional com curadoria, descrições estruturadas (metadados) dos dados publicados, para fornecer uma nova estrutura de compartilhamento de dados e de reutilização que, acreditamos, irá acelerar o ritmo das descobertas científicas (NATURE, 2019, *tradução nossa*).

Segundo Aventurier e Alencar (2016), desde 2010 estão disponíveis periódicos que publicam artigos descrevendo os conjuntos de dados, as condições de produção e o interesse científico, os chamados *data papers* (artigo de dados). Este tipo de periódico se fundamenta em seis princípios: **crédito** – dar reconhecimento aos cientistas que elaboram conjuntos de dados disponíveis e reutilizáveis; **reúso** – padronizações e descrições detalhadas tornam os dados de pesquisa mais fáceis de se encontrar e reutilizar; **qualidade** – avaliação crítica para verificar o rigor experimental e a totalidade de sua descrição; **descoberta** – facilidade em localizar os conjuntos de dados que são relevantes para suas pesquisas; **aberto** – os cientistas trabalham melhor quando podem facilmente se conectar e colaborar com seus pares; **serviço** – o periódico se compromete a fornecer um serviço de excelência tanto para autores quanto para leitores (NATURE, 2019, *tradução nossa*).

Diferentemente dos periódicos tradicionais, os *data journals*, conforme afirmam Curty e Aventurier (2017, p. 15), não apresentam seções de revisão de literatura, análise de dados e conclusão. Todavia, são estruturados com tópicos de descrição dos metadados, como: resumo, introdução, contexto de coleta, agradecimentos, referências e, em alguns casos, informam se há conflito de interesses. Para os autores, os *data journals* têm mais chance de serem aceitos pela comunidade científica, pois mantêm os ritos, observados nos periódicos tradicionais, como a revisão por pares e a periodicidade.

A criação dos *data papers* demonstra a preocupação dos setores ligados ao avanço científico com a disponibilização dos dados, oriundos de pesquisas científicas, e a preservação deles pelo maior tempo possível. Nesse sentido, a investigação realizada por pesquisadores do Canadá avaliou a preservação dos dados de artigos publicados pelos autores e constatou que:

[...] a perda de dados é maior quanto mais antiga é a publicação. A probabilidade de que os dados de um artigo tenham sido conservados pelo autor cai de um fator de 17% ao ano. Soma-se a isso a dificuldade de localizar os autores, uma vez que publicações mais antigas não incluem endereços eletrônicos ou estão desatualizados. A probabilidade de contatar um autor cai na razão de 7% ao ano. Desta forma, estima-se que 80% dos dados não deverão estar disponíveis decorridos 20 anos de sua geração (SCIENTIFIC ELECTRONIC LIBRARY ONLINE, 2016, *on-line*).

Além da criação dos periódicos de dados, outra iniciativa são as políticas de compartilhamento. O estudo desenvolvido por Stodden, Guo e Ma (2013) investiga 170 políticas de compartilhamento de dados de periódicos da área da Ciência da Computação, no período de junho de 2011 a junho de 2012. Os autores descobriram que a maioria dos periódicos científicos da área não exigia o depósito de dados; apenas o encorajava.

No contexto amplo da temática dos periódicos científicos, outro marco da história é a pesquisa realizada por Larivière, Haustein e Mongeon (2015). Esses pesquisadores afirmam que a trajetória dos recursos é marcada pela consolidação do oligopólio de seis grandes editoras comerciais. O estudo analisou 45 milhões de documentos indexados na plataforma *Web of Science*, entre 1973 e 2013, e observou o crescimento de seis editoras comerciais: *Elsevier, Blackwell, Springer, Taylor & Francis, American Chemical Society e Sage*. Esse grupo concentra 70% da produção em ciências naturais, ciências médicas e ciências sociais aplicadas indexadas na *Web of Science*. Os autores comentam o crescimento, ressaltando que, por exemplo, em 1995 a concentração dos artigos de ciências sociais nessas editoras era de 15%, passando, em 2013, para 66% das publicações, fato este que destaca a comercialização e o monopólio do conhecimento científico.

Para Morrison (2012), segundo consta em sua tese de doutorado, as editoras comerciais possuem uma receita lucrativa. Neste estudo, a autora apresenta dois fatos. O primeiro é rentabilidade da comercialização de periódicos científicos; e o segundo é que, apesar das alternativas propostas pelo Movimento de Acesso Aberto, o mercado tem se mostrado em crescimento. Ainda conforme a autora, as editoras comerciais alcançaram altas taxas de lucro – entre 32% e 42% – que crescem, ao longo dos anos, entre 3% e 13%. Para Costa (2017), a lucratividade da comercialização da comunicação científica foi um ponto fortemente questionado no início dos anos 2000 nas articulações do Movimento de Acesso Aberto, o qual propõe uma alternativa a esse modelo.

A constituição do Movimento de Acesso Aberto se deu a partir do objetivo e algumas iniciativas voltadas à disponibilização, de forma livre e irrestrita, dos resultados das pesquisas científicas em texto completo por meio da Internet (LYNCH, 2003). Para Costa (2017), 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”. O chamado Acesso Aberto (*Open Access*) surgiu como solução às dificuldades enfrentadas com o cenário de monopólio das editoras sobre as publicações científicas. Pois

tal contexto limita o acesso aos resultados das pesquisas a pesquisadores que estão ligados a instituições que adquirem o acesso à informação. Segundo Sales e Sayão (2012), “esse fato impede que o conhecimento científico – um patrimônio da humanidade – circule livremente e esteja disponível para todos”. Para Costa (2017), as proposições e iniciativas do acesso aberto pertencem à filosofia aberta.² No campo científico, os desdobramentos da filosofia aberta provocaram discussões sobre a Ciência Aberta.

2.1.3 Ciência Aberta

Os princípios que sustentam a ideia da Ciência Aberta, de acordo com Boulton (2013), são os mesmos que serviram de base para as revoluções científicas dos séculos XVIII e XIX, porque, em ambos os contextos, a motivação seria tornar a ciência pública. Para 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*.³

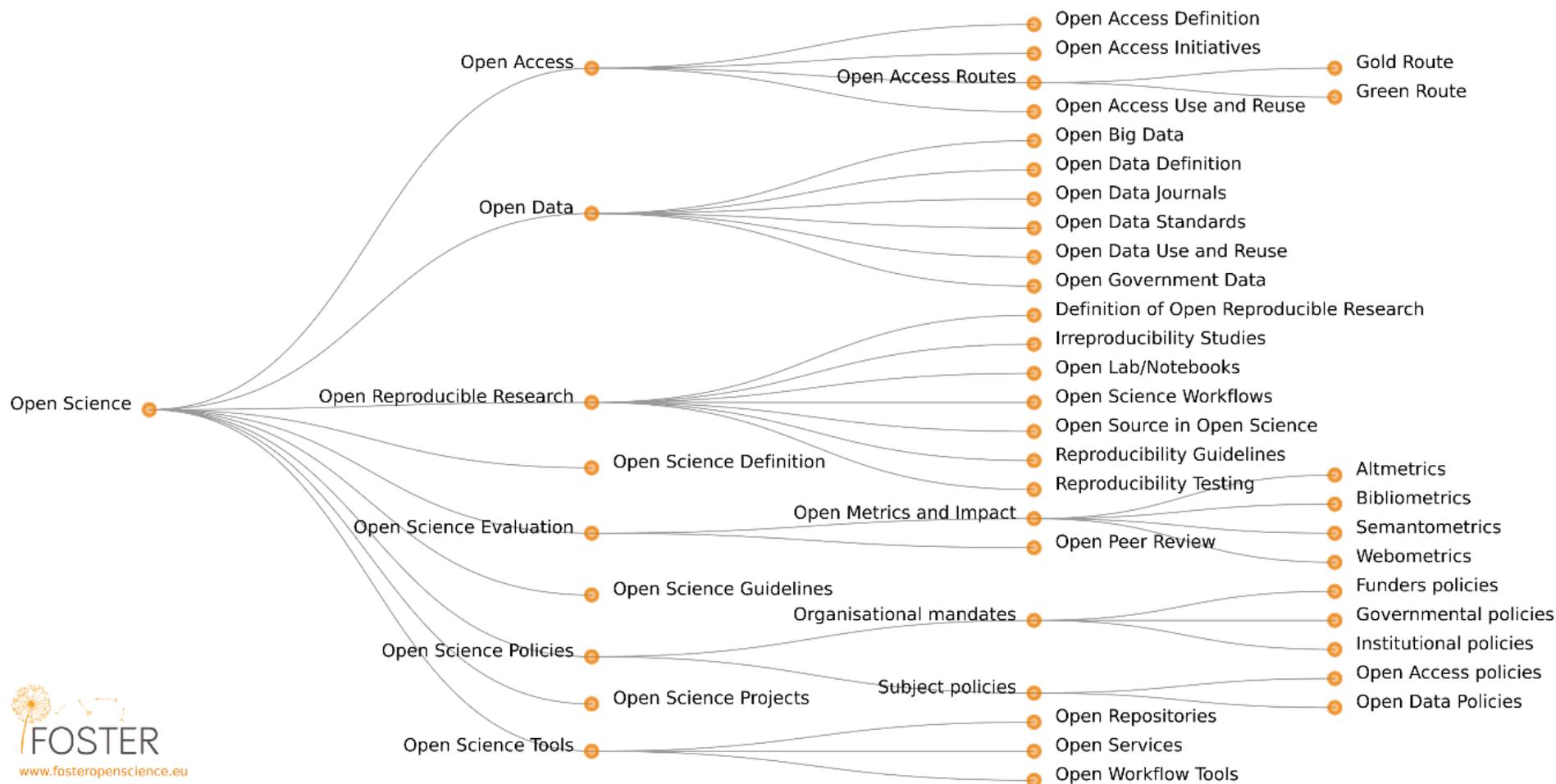
O movimento Ciência Aberta provoca mudanças na forma de conduzir a ciência, inserindo novos conjuntos de práticas sobre o uso da internet e das ferramentas da web social (ou redes sociais), impactando desde a formulação das questões e das hipóteses até a difusão dos resultados de pesquisas, ou seja, implicando em uma mudança na abordagem acadêmica. A Ciência Aberta constitui um termo guarda-chuva, o qual agrega práticas e abordagens (ALBAGLI, 2015, p. 15). Para Pontika *et al.* (2015), o projeto *Facilitate Open Science Training For European Research* (FOSTER) indica seis categorias para se pensar a Ciência Aberta (Figura 6): *Open Access, Open Data, Open Reproducible Research, Open Science Evaluation, Open Science Policies, Open Science Tools*.

² Segundo Costa (2006), a filosofia aberta é um conjunto de ideias que “denotam um novo modelo de representar um igualmente novo processo de comunicação científica, ao mesmo tempo em que serve de base para interpretá-lo”.

³ Capacidade de autocorreção da ciência.

Figura 6 - Taxonomia da Ciência Aberta

Open Science Taxonomy



As seis categorias indicadas pelo projeto FOSTER são discutidas por Costa (2017, p. 36) em sua tese (Quadro 1). Para a autora, as categorias podem ser descritas da seguinte maneira: a primeira delas, *Open Access*, aborda a definição e as iniciativas do Movimento de Acesso Aberto, além de suas estratégias, que são a via verde e a dourada. A segunda categoria, *Open Data*, trata dos dados abertos que incluem ampla natureza, desde dados de pesquisa até dados de governo. A terceira, *Open Reproducible Research*, discute a reprodutibilidade da pesquisa. A quarta categoria, *Open Science Evaluation*, diz respeito às formas de avaliação da ciência, que devem considerar métricas abertas e qualificar seu impacto em sentido amplo. A quinta categoria, *Open Science Policies*, ocupa-se das políticas da Ciência Aberta. Por fim, a última categoria, *Open Science Tools*, engloba as ferramentas que serão utilizadas na pesquisa.

Quadro 1 - Categorias para se pensar a Ciência Aberta

	CATEGORIAS	DEFINIÇÃO
1	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.
2	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.
3	Metodologias abertas	A pesquisa deve ser desenvolvida a partir de métodos e processos consistentes. Isso permitirá que ela possa ser reproduzida ou analisada corretamente.
4	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 webometria.
5	Políticas abertas	São os documentos das instituições ou agências fomentadoras de pesquisa que promovem as iniciativas de acesso aberto nas pesquisas que financiam.
6	Ferramentas abertas	A pesquisa deverá apoiar-se em tecnologias abertas, tanto softwares como hardwares livres.

Fonte: (COSTA, 2017, p. 36-37).

Moreno (2018) relata que, no contexto europeu, o projeto FOSTER trabalha com a proposta de auxiliar os diferentes membros da comunidade científica (estudantes, pesquisadores, entre outros) a promover a ciência de acesso aberto. Para tanto, o projeto define *Open Science* como “[...] o movimento para tornar a pesquisa, os dados e a disseminação científica acessíveis a todos os níveis de uma sociedade investigativa” (FOSTER, 2019, *on-line*).

Sob essa perspectiva, como defendem Fecher e Friesike (2014), a Ciência Aberta pode significar um conjunto de conceitos “que vão desde o democrático direito de acesso ao conhecimento (por exemplo, acesso aberto a publicações), até a demanda por incluir o público na pesquisa (por exemplo, ciência cidadã) para o uso de ferramentas de colaboração e de compartilhamento”.

O conceito de dados abertos vem sendo discutido ao longo do tempo, uma vez que a partir de sua definição é possível que os movimentos que defendem a abertura do conhecimento elaborem planos de ação. Nesse caso, a *Open Knowledge Foundation* (OKF) desenvolveu uma definição bem abrangente, composta por onze itens (Quadro 2) que envolvem aspectos técnicos, legais e procedimentais no uso e na distribuição da informação (OPEN KNOWLEDGE FOUNDATION, 2019).

Quadro 2 - Itens para ser considerados Open, segundo a OKF

1.	Acesso	A obra deve ser disponibilizada na íntegra, por um preço que não exceda o custo razoável de reprodução; preferencialmente gratuita na Internet. A obra também deve estar disponível em uma forma utilizável e modificável.
2.	Redistribuição	A licença não deve restringir a possibilidade de venda ou distribuição da obra em si ou como parte de um pacote com obras de fontes diversas. A licença não deve exigir pagamento de direitos ou outra taxa para venda ou distribuição.
3.	Reutilização	A licença deve permitir modificações e obras derivadas; deve permitir que estas sejam distribuídas sob as mesmas condições da obra original.
4.	Ausências de restrições tecnológicas	Não deve haver restrições tecnológicas. A disponibilização da obra deve estar em formato cuja especificação seja disponível pública e gratuitamente, e cujo uso não esteja sujeito a restrições monetárias ou outras.
5.	Atribuição	Como condição para a redistribuição e a reutilização, a licença pode exigir atribuição da autoria – desde que de forma não onerosa – aos autores da obra.
6.	Integridade	É aceitável que, como condição para a distribuição da obra, se exija, no caso de sua modificação, nome ou número de versão diferente daquele da obra original.
7.	Não discriminação de pessoas ou grupos	A licença não discriminará indivíduos ou grupos de indivíduos.
8.	Não discriminação de domínios de atividade	A licença não pode restringir o uso da obra a uma área de atividade específica.
9.	Distribuição da Licença	Os direitos da obra devem ser aplicados na dela, sem necessidade de uma licença adicional.

10.	A Licença não deve ser específica de um pacote	Os direitos à obra não devem depender de sua inserção em um determinado pacote. Cada obra deve ter os mesmos direitos que o conjunto do pacote.
11.	A Licença não deve restringir a distribuição de outras obras	A licença não deve restringir outras obras que sejam distribuídas conjuntamente com a obra licenciada. Por exemplo, a licença não deve impor que todas as obras distribuídas pelo mesmo meio sejam abertas.

Fonte: (MACHADO, 2015).

No contexto da Ciência Aberta, os dados de pesquisa estão no centro de uma nova discussão, visto que por muito tempo os cientistas se preocupavam apenas em disseminar os seus resultados de pesquisa, como nos modelos de comunicação científica propostos por Garvey e Griffith (1972), Hurd (2000) e Björk (2007). Mas conforme argumenta Molloy (2011), comunicar os dados de pesquisa melhora a ciência em termos de transparência, reprodutibilidade, eficiência e, em última instância, traz maior benefício para a sociedade.

Ao longo do tempo, os modelos de comunicação se modificaram, já que novas tendências e ferramentas surgiram para ajudar os processos da comunicação científica. Conseqüentemente, o compartilhamento de dados – que é uma prática que vem sendo aceita pela comunidade científica – reaparece no fluxo da comunicação científica (BJÖRK, 2007; UNESCO, 1971). Tal prática vem ganhando espaço, uma vez que compartilhar informação sempre contribuiu com o progresso da ciência. Além do mais, compartilhar dados de pesquisa com a comunidade permite gerar mais análises sob diferentes perspectivas, confirmar os resultados e explorar novas hipóteses.

Os estudos sobre os dados discorrem sobre a importância do profissional da informação na gestão de dados de pesquisa. Isto ocorre, porque os dados de pesquisa impactam na comunicação científica, a partir do momento em que se necessita dar destaque aos processos intermediários da pesquisa, por meio, da criação e da gestão de repositórios de dados, dos padrões de metadados e dos planos de gestão da pesquisa, este último muitas vezes são necessários para o próprio financiamento da pesquisa (ALLARD; MACK; FELTNER-REICHERT, 2005; CHOUDHURY, 2008; STEINHART, 2006).

De acordo com Costa (2017), a comunicação dos dados de pesquisa tem se mostrado relevante nas discussões sobre a Ciência Aberta. Porém, na literatura não há uma definição sobre o conceito de dado de pesquisa, visto que o entendimento pode variar de acordo com a área de conhecimento e os próprios pesquisadores (GUIBAULT; WIEBE, 2013; SAYÃO; SALES, 2015).

2.2 Comunicação dos dados de pesquisa

2.2.1 Dados de pesquisa

De acordo com o relatório da *National Academy of Science*, os dados de pesquisa, ou dados científicos, “são fatos, números, letras e símbolos que descrevem um objeto, uma ideia, uma condição, uma situação ou outros fatores”. O estudo realizado pela equipe da biblioteca do *Georgia Institute of Technology*⁴ (Instituto de Tecnologia da Geórgia) apresenta a seguinte definição para dados de pesquisa:

[...] informação digital estruturada por metodologia formal com o objetivo de criar novas pesquisas ou bolsa de estudos. Esses dados podem estar em uma variedade de formatos adequados para comunicação, interpretação ou processamento, incluindo leituras de sensores, pesquisas e equipamentos de laboratório, modelos de simulação e bancos de dados compilados e arquivos de texto, entre outros. (PARHAM; BODNAR; FUCHS, 2012, p. 10, *tradução nossa*).

Em termos de definições, Torres-Salinas, Robinson-Garcia e Cabezas-Clavijo (2012, p. 175) afirmam que a de maior consenso é a que considera os dados de pesquisa como todo o material registrado durante a pesquisa, reconhecido pela comunidade científica, e que serve para certificar os resultados da pesquisa realizada. Os autores dizem que essa definição é aceita pelo NIH e *Organisation for Economic Cooperation and Development* (OECD).

Guibault e Wiebe (2013) afirmam que dados de pesquisa são “os dados produzidos no decorrer da pesquisa”. Já a OECD (OECD, 2007) e Padilla Navarro *et al.* (2013) destacam que dados de pesquisa são utilizados como fonte primária em uma pesquisa, não necessariamente produzidos pela própria pesquisa. Rodrigues *et al.* (2010) ampliam o escopo da definição, incluindo “todos os dados usados de algum modo na pesquisa”. De acordo com o *Guidelines on open access to scientific publications and research data in Horizon 2020* (2016), os dados de pesquisa referem-se à informação e, em particular, aos fatos ou números coletados que são examinados e considerados como base para argumentação, discussão ou cálculo.

Ao considerar as perspectivas aqui apresentadas, a definição proposta por Costa (2017) sintetiza todas as anteriores e é considerada a mais aproximada para este estudo. Aliás, a autora entende que “os dados de pesquisa são os dados produzidos como fontes primárias,

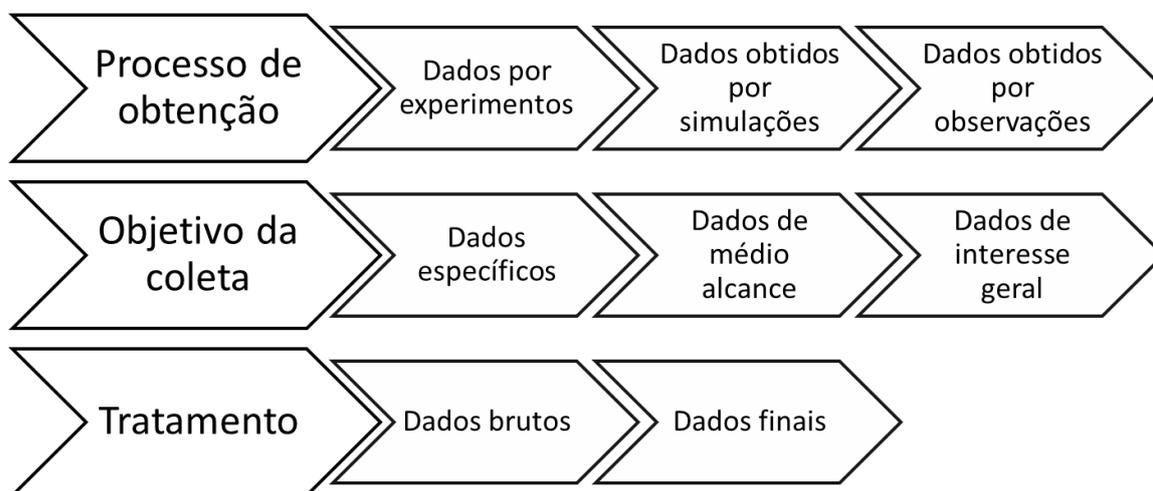
⁴ Possui seis faculdades e 28 escolas com foco em administração, computação, design, engenharia, artes e ciências.

ou utilizados para o desenvolvimento de uma pesquisa, os quais têm por finalidade embasar a argumentação da pesquisa”.

A diversidade dos dados de pesquisa ocorre pelo fato de serem gerados com diferentes propósitos. Diversas disciplinas têm e usam linguagem disciplinar específica em torno dos dados de pesquisa (HENDERSON, 2017; KELLAM; THOMPSON, 2016; KOLTAY, 2017; NATIONAL..., 1999; PIORUN, 2013; SAYÃO; SALES, 2016). Para Semeler e Pinto (2019), os dados de pesquisa abrangem diferentes disciplinas acadêmicas. Eles representam conjuntos de dados que estão registrados nas mais diversas áreas do conhecimento e que estão distribuídos globalmente.

Além da questão da definição do conceito, há algumas tentativas de classificar os dados de pesquisa. De acordo com Torres-Salinas, Robinson-Garcia e Cabezas-Clavijo (2012), uma das classificações mais elementares é a que se refere aos próprios formatos dos dados (texto, imagens, vídeo etc.). Em contrapartida, a classificação oferecida pela *Research Information Network* (RIN), apresentada por Swan e Brown (2008), indica três critérios para classificar os dados de pesquisa. São eles: processo de obtenção, objetivo da coleta e tratamento.

Figura 7 - Classificação dos dados de pesquisa



Fonte: Elaboração própria baseada em Swan e Brown (2008).

O primeiro critério estabelece três tipos de dados: experimentais, simulações e observações. Os dados experimentais seriam aqueles obtidos como resultado de um experimento normalmente reproduzível. Por simulações entenderíamos todos os dados gerados como resultado da aplicação de um modelo. Finalmente, os dados de observação

seriam o resultado da observação direta de um fenômeno em tempo real e são únicos e insubstituíveis como os resultados de uma pesquisa.

O segundo critério também se divide em três tipos. Os dados específicos são aqueles que respondem às necessidades de um projeto de pesquisa em especial, cujo valor é muito limitado. A seguir, surgem os dados de médio alcance, que são aqueles que oferecem benefício a apenas uma comunidade específica. E, finalmente, dados de interesse geral, os quais seriam aqueles cuja importância é vital para o desenvolvimento da ciência e que despertam grande interesse em toda a comunidade científica.

O terceiro critério tem relação com a fase de pesquisa, ou seja, há uma distinção entre dados preliminares e dados finais. O primeiro seria o recém-extraído, que não recebeu nenhum tipo de tratamento do pesquisador (dados brutos). Os dados finais correspondem ao que o NIH também define como dados finais da pesquisa e que são o resultado da combinação ou processamento dos dados preliminares.

Em 2007, Lyon cita duas classificações para dados. A primeira do *Natural Environment Research Council* (NERC),⁵ que divide os dados em: dados canônicos (que não mudam) e dados episódicos (que podem mudar). Já a *International Union of Crystallography* (IURC) classifica os dados em três tipos: dados crus (imagem de arquivos), dados primários (fatores de estrutura) e dados derivados (seis dimensões do modelo estrutural).

Harvey (2010) apresenta outra classificação para os dados. O autor os divide em coleções: coleção de pesquisa, formada por dados locais gerados em um laboratório ou projeto; coleção de comunidade, formada por bases de dados internacionais geradas a partir de pesquisas integradas; e coleção de referência, que reúne dados referenciais sobre determinados experimentos.

De acordo com Borgman (2012, p. 3), outra maneira de classificar os dados é atribuindo valor a eles. Nesse sentido, a autora discorre que a valoração dos dados de pesquisa é complexa, pois depende do contexto em que eles estão inseridos, de modo que os mesmos podem ter valores tanto imediatos quanto duradouros. Ao mesmo tempo que alguns têm valor momentâneo, outros, no entanto, adquirem-no ao longo do tempo. E, além de todos esses, existem os dados que podem ser facilmente recriados.

⁵ Conselho de pesquisa britânico, que apoia as atividades de pesquisa, formação e transferência de conhecimento nas ciências ambientais.

Conforme a *National Science Foundation* (NSF), em relatório sobre Coleções de Dados Digitais de Longa Vida, publicado em 2007, os dados também podem ser classificados de acordo com a natureza, a origem e o nível de processamento ao qual foram submetidos. Em relação à natureza, os dados de uma coleção podem ser variados – por exemplo, informações de versão de software, números, imagens, algoritmos etc. Quanto às origens, os dados podem ser observacionais, computacionais ou experimentais (Quadro 3). Tal distinção é fundamental para as escolhas destinadas a arquivamento e preservação.

Quadro 3 - Tipo de dados em relação à origem

TIPO DE DADO	DESCRIÇÃO
Dados observacionais	Gerados a partir de observações diretas, por exemplo: a temperatura do oceano em uma data específica, a atitude dos eleitores antes de uma eleição, as fotografias de uma supernova ou, ainda, os dados gerados a partir de um acidente nuclear.
Dados computacionais	Originados a partir da execução de um modelo de computador ou simulação. Inclui entre eles as informações sobre o modelo (somado a uma descrição completa do hardware, software e dados de entrada), bem como sua disponibilidade.
Dados experimentais	Originados a partir de experiências, tais como medições de padrões de expressão genética, taxas de reação química ou desempenho de um motor.
Dados de registros	Gerados por empresas públicas ou privadas e que são úteis para a pesquisa científica, social e humanística.

Fonte: Adaptado de NSF (2007).

Borgman (2010) defende, especialmente, duas formas de classificar os dados, formas essas baseadas nos estágios de geração dos dados: dados brutos e dados derivados. De acordo com a autora, os dados brutos são aqueles recolhidos inicialmente na sua forma bruta e, dependendo dos objetivos da pesquisa, sujeitos a estágios posteriores de refinamento e análise. Os dados derivados são gerados a partir de atividades de processamento ou curadoria.

Importa ressaltar que as classificações foram construídas em contextos específicos, por isso diferindo umas das outras, apesar de atenderem aos objetivos pelos quais foram criadas. Tal afirmação é embasada pelo estudo de Souza (2012, p. 4), o qual afirma que “os mesmos objetos e ideias podem ser organizados e representados de formas diferentes e toda classificação está relacionada a um propósito definido de construção e uso de informação”. Portanto, as áreas de conhecimento são as responsáveis por estabelecer uma classificação para dados de pesquisa, pois devem-se considerar os métodos de pesquisa que os geram e as práticas e normas da área (TORRES-SALINAS; ROBINSON-GARCIA; CABEZAS-CLAVIJO, 2012).

Quadro 4 - Síntese dos tipos de dados de pesquisa

CARACTERÍSTICA DE DIVISÃO	SWAN; BROWN	LYON	HARVEY	BORGMAN	NSF
Quanto à natureza					Número Imagem Software
Quanto à origem	Experimental Simulações Observações			Registro	Observacionais Computacionais Experimentais
Quanto ao nível de processamento e/ou estágio de geração	Preliminares Finais	Crus Primários Derivados		Brutos Derivados	Intermediário Finais
Quanto à formação de coleções			De pesquisa De comunidade De referência		
Quanto à mutabilidade		Canônicos Episódicos			
Quanto ao valor	Específicos Médio alcance De interesse geral				

Fonte: Elaborado a partir das definições de Borgman (2010), Harvey (2010), Lyon (2007), National Science Foundation (2007) e Swan e Brown (2008).

Para a comunicação dos dados de pesquisa, é necessário que esses estejam descritos, para que depois possam ser encontráveis, acessíveis, interoperáveis e reutilizáveis. Assim, as próximas duas subseções tratam das definições de metadados e dos princípios FAIR (do inglês FAIR – Findable, Accessible, Interoperable, and Reusable).

2.2.2 Metadados

De acordo com Portal Brasileiro de Dados Abertos, “metadados são dados sobre os dados, ou seja, são informações que possibilitam organizar, classificar, relacionar e inferir novos dados sobre o conjunto de dados” (BRASIL, 2019). Na perspectiva da Ciência da Informação, segundo afirmam Arakaki e Arakaki (2019, p. 61), o conceito de metadados está atribuído a uma informação estruturada para as ações de identificação, descoberta, seleção, uso, acesso e gerenciamento. Para Gilliland (2016), o termo metadado é ainda compreendido

de diversas maneiras pelas diferentes comunidades e profissionais que desenvolvem, criam, descrevem, preservam e usam sistemas e recursos de informação.

No que se refere às funções dos metadados, identifica-se que estão atreladas a sua tipologia (GILLILAND, 2016). O estudo desenvolvido por Arakaki (2019, p. 80–81) identifica as 10 tipologias de metadados, que são: administrativos; autenticação; preservação; proveniência; técnicos; *Metametadata*; descritivos; direitos; acesso e uso; estruturais; e *Markup languages* (Linguagem de marcação) (Quadro 7).

Quadro 5 - As tipologias dos metadados e suas definições

TIPOLOGIA	DESCRIÇÃO
Metadados administrativos	usados para gerenciar e administrar coleções e recursos informacionais, para auxiliar na tomada de decisão e manutenção dos registros e recursos informacionais. Fornecem informações sobre a origem e a manutenção de um objeto
Metadados de autenticação	são informações que possibilitam a identificação, integridade, legitimidade de um recurso informacional
Metadados de preservação	estão relacionados com informações de preservação e conservação dos recursos informacionais
Metadados de proveniência	estão relacionadas às informações de procedência, fornece dados sobre entidades, criação e modificações e seus relacionamentos
Metadados técnicos	estão relacionados a como um sistema funciona, fornecendo informações do sistema ou do recurso
<i>Metametadata</i>	estão relacionados a como um sistema funciona, fornecendo informações do sistema ou do recurso
Metadados descritivos	identificam características e contextos intelectuais dos recursos de informação para fins de descoberta, identificação, seleção, aquisição, contexto e compreensão
Metadados de direitos	identificam características e contextos intelectuais dos recursos de informação para fins de descoberta, identificação, seleção, aquisição, contexto e compreensão
Metadados de acesso e uso	são informações de como um recurso informacional foi acessado e usado, como restrições de circulação e acesso, registros de exposições, entre outros
Metadados estruturais	está relacionado à composição e organização do recurso informacional
<i>Markup languages</i>	integra metadados e sinalizações para outros recursos estruturais ou semânticos

Fonte: Arakaki e Arakaki (2019, p. 62–63).

Ainda é perceptível a relação das tipologias dos metadados com os padrões de metadados adotados em sistemas de informação (ZENG; QIN, 2016). No que diz respeito aos padrões de metadados, a adoção varia de acordo com a área do conhecimento, de maneira

que disciplinas, repositórios ou centros de dados específicos podem orientar ou mesmo estabelecer o conteúdo e o formato dos metadados, possivelmente usando um padrão formal (OPENAIRE, 2018, *on-line*). Nesse sentido, como afirma *Digital Curation Centre* (DCC) (2019, *on-line*), há padrões de metadados gerais e específicos, como, por exemplo:

- a) Dublin Core – padrão de metadados agnóstico de domínio, básico e amplamente usado;
- b) DDI (*Data Documentation Initiative*) – padrão comum para ciências sociais, comportamentais e econômicas, incluindo dados de pesquisa;
- c) EML (*Ecological Metadata Language*) – específico para disciplinas de ecologia;
- d) ISO 19115 e FGDC-CSDGM (*Federal Geographic Data Committee's Content Standard for Digital Geospatial Metadata*) – para descrever informações geoespaciais;
- e) MINSEQE (*MINimal information about high throughput SEQuencing Experiments*) – padrão genômico;
- f) FITS (*Flexible Image Transport System*) – padrão de arquivo digital de astronomia que inclui metadados estruturados e incorporados;
- g) MIBBI (*Minimum Information for Biological and Biomedical Investigations*) – Informações Mínimas para Investigações Biológicas e Biomédicas;

O *Dublin Core Metadata*, conforme afirma Sales (2014), é o padrão de metadados mais difundido atualmente. Surgiu em 1995, em um Workshop promovido pela *Online Computer Library Center* (OCLC), e sua aplicação vem sendo disseminada para representação de diversos tipos de objetos digitais. Para Santos, Simionato e Arakaki (2014, p. 149) os metadados influenciam tanto no acesso quanto na recuperação da informação, mas para que tenham aceitação internacional e auxiliem na interoperabilidade eles devem estar padronizados. De forma que, como afirma Méndez Rodríguez (2001, p. 787) o padrão de metadados é como um conjunto de elementos que compõem uma especificação (geralmente associada a nomes de um domínio de informações) dos rótulos que devem ser usados com um valor dêitico para um objeto de informação. Segundo a autora, mesmo sendo criados para o contexto tecnológico, muitos dos padrões de metadados desenvolvidos não atendem satisfatoriamente às necessidades de representação informacional em um domínio específico, porque os esquemas de descrição são generalizantes.

Na perspectiva da comunicação dos dados, o gerenciamento de metadados é necessário para tornar os dados disponíveis, utilizáveis, compartilháveis e armazenáveis

(MAYERNIK, 2011). Para o autor, os pesquisadores em muitos ambientes científicos têm pouca experiência ou conhecimento em gerenciamento de dados e metadados. Ao contrário do que acontece com as publicações, conforme afirmam Gómez, Méndez e Hernández-Pérezon (2016), que apesar das diferenças entre as áreas do conhecimento há um núcleo comum de propriedades formais. Contudo, para os autores, os dados científicos mostram uma heterogeneidade que varia radicalmente entre disciplinas, áreas temáticas e mesmo entre grupos de pesquisa e pesquisadores. De forma que, a construção de padrões de metadados, que atendam as necessidades de cada área do conhecimento, se faz importante para o compartilhamento de dados. Visto que, como afirma Borgman (2010), sem os metadados e as descrições da metodologia e do contexto da pesquisa, os dados são apenas coleções de números, listas de códigos, belas fotos ou caixas de pedras.

Um padrão de metadados define parâmetros que devem ser utilizados de forma global, tornando mais prático e seguro o compartilhamento dos conjuntos de dados (SILVA *et al.*, 2015, p. 29). Para os autores, os padrões de metadados estabelecem regras consistentes e vocabulários padronizados, que possibilitam eficácia e eficiência na comunicação entre os geradores e consumidores de dados. Dessa forma, quanto mais adequado o registro de metadados melhores serão as condições para que os dados sejam validados, registrados, preservados, disseminados e reutilizados por outros pesquisadores.

Nesse sentido, no contexto da Ciência Aberta, foram pensados princípios para nortear as boas práticas para publicação dos dados científicos, chamados FAIR (do inglês FAIR – *Findable, Accessible, Interoperable, and Reusable*)⁶.

2.2.3 Os princípios FAIR (do inglês FAIR – *Findable, Accessible, Interoperable, and Reusable*)

Em 2014, na conferência internacional denominada *Jointly designing a data FAIRPORT*, criaram-se os princípios FAIR para discutir a utilização, o tratamento e o reúso dos dados de pesquisa no âmbito *e-Science* com especialistas de diversas áreas do conhecimento (HENNING; RIBEIRO; SALES; *et al.*, 2019).

O termo *e-Science*⁷ foi empregado em 2001 por John Taylor, enquanto diretor geral do Conselho de Pesquisa do *Office of Science and Technology* (OST) do Reino Unido, para

⁶ Em português: Encontráveis, Acessíveis, Interoperáveis e Reusáveis.

⁷ A literatura revela que a *e-Science*, além de ter grafias diferenciadas como eScience, escience, EScience etc., também pode receber o nome de Cyberinfrastructure, cyberscience, eInfrastructure e eResearch (MEDEIROS; CAREGNATO, 2012, p. 315). Para esta pesquisa, optou-se pelo uso do termo *e-Science*.

denominar novas formas de colaboração, trabalho multidisciplinar em áreas-chave da ciência e a infraestrutura necessária que permite essas novas formas de trabalho (HEY; TREFETHEN, 2002). Dentre as definições que se encontram na literatura sobre *e-Science*, Hey e Hey (2006, p. 517, *tradução nossa*) a entendem como um conjunto de ferramentas e tecnologias que devem apoiar a ciência em rede. Ribes e Lee (2010, p. 232) assinalam que a *e-Science* é um nome dado “para as tecnologias de informação em rede de apoio a atividades de investigação científica, como a colaboração de compartilhamento de dados e divulgação dos resultados”.

O Guia dos Princípios FAIR foi criado em 15 de março de 2016 com o objetivo de contribuir com a criação de boas práticas para a publicação de dados científicos que sejam amplamente compartilhadas, claramente articuladas e amplamente aplicáveis (SCIENTIFIC ELECTRONIC LIBRARY ONLINE, 2016). Os princípios (Quadro 5) estabelecem orientações estruturais para a publicação de recursos digitais, conjunto de dados, códigos e objetos de pesquisa. São considerados o instrumento norteador para a descoberta, o acesso, a interoperabilidade, o compartilhamento e a reutilização dos dados de pesquisa (WILKINSON, Mark D. *et al.*, 2016).

Quadro 6 - Guia dos princípios FAIR

Princípios	Como aplicar
F – <i>Findable</i>	Para serem encontrados
F1. Os (meta)dados devem ter identificadores globais, persistentes e identificáveis	Metadados com identificador único - <i>Unique and persistent identifier</i> (PID) ou equivalente, para cada conjunto de dados. P. Ex.: DOI, ARK, RRID, PID.
F2. Os dados devem ser descritos com metadados enriquecidos (impacta diretamente R1)	Conjuntos de dados devem possuir metadados ricos o suficiente para que, uma vez indexados para um mecanismo de busca, esses metadados possam ajudar o usuário dos dados a encontrá-los mesmo que não possua o seu identificador.
F3. Os metadados devem incluir claramente e explicitamente os identificadores dos dados que descrevem	Como não podemos prever que os dados e seus metadados estejam sempre juntos (por exemplo, quando os metadados são indexados por um mecanismo de busca e, portanto, estão em uma plataforma diferente dos dados), a associação entre eles deve ser feita pela inclusão do identificador dos dados pelos metadados.
F4. Os (meta)dados devem ser registrados ou indexados em recursos que ofereçam capacidades de busca	A fim de contribuir para que os dados sejam encontrados, seus metadados devem ser indexados por mecanismos de busca que, por sua vez, permitem aos usuários encontrá-los por meio de elementos desses metadados.
A – <i>Accessible</i>	Para serem acessíveis
A1. (Meta)dados devem ser recuperáveis pelos seus identificadores usando protocolo de comunicação padronizado	Com o identificador do conjunto de dados ou de seus metadados, o usuário deve poder recuperá-los por meio de um protocolo de comunicação padronizado.

A1.1 O protocolo deve ser aberto, gratuito e universalmente implementável	Independente de licenciamento dos dados e dos metadados, o protocolo de comunicação usado para dar acesso a eles deve ser aberto, gratuito e passível de ser implementado por qualquer interessado.
A1.2 O protocolo deve permitir procedimentos de autenticação e autorização, quando necessário	Quando necessário, dependendo das restrições ao acesso aos dados e/ou metadados, um mecanismo de autenticação e autorização para o acesso deve ser permitido pelo protocolo de comunicação.
A2. Metadados devem ser acessíveis, mesmo quando os dados não estão mais disponíveis	É preciso existir um conjunto de estratégias de preservação para os dados e metadados. Minimamente, os metadados devem ser sempre acessíveis, possibilitando a criação de índices para conjuntos de dados atuais e passados.
<i>I – Interoperable</i>	Para serem interoperáveis.
I1. (Meta)dados devem ser representados por meio de uma linguagem formal, acessível, compartilhada e amplamente aplicável para a representação do conhecimento	Para que possamos representar dados e metadados devemos utilizar linguagens de representação do conhecimento que sejam formais, acessíveis e amplamente aplicáveis. Por exemplo, RDF, XML, DICOM etc.
I2. (Meta)dados devem usar vocabulários que seguem os princípios FAIR	Como os dados e metadados devem possuir referências aos vocabulários que contenham os conceitos utilizados, devemos garantir que sejam utilizados vocabulários que também sigam os princípios FAIR.
I3. (Meta)dados devem incluir referências qualificadas para outros (Meta) dados	Referenciar os conjuntos de dados devidamente, possibilitando que conjuntos de dados gerados, a partir de outros conjuntos de dados, sejam ligados. Possibilitar a ligação semântica entre eles por meio dos padrões adotados.
<i>R – Reusable</i>	Para serem reutilizáveis
R1. (Meta)dados são descritos com uma pluralidade de atributos precisos e relevantes.	Prover metadados que permitam aos potenciais usuários avaliarem se é possível reusar os dados ou metadados e também, se eles se adequam às suas necessidades.
R1.1. (Meta)dados devem ser disponibilizados com licenças de uso claras e acessíveis	É fundamental que o responsável pelos dados e metadados defina explicitamente quem pode ter acesso a eles, para que e sob quais condições. Essas informações são definidas por meio de suas licenças de uso.
R1.2. (Meta)dados devem estar associados à sua proveniência	Especificar a proveniência (linhagem) dos dados é importante não só para que o usuário possa avaliar a utilidade dos dados ou metadados, mas também para que possamos atribuir o devido crédito a quem produziu, manteve ou editou esses dados. Dentre as informações relevantes à proveniência destacam-se: (a) A linhagem dos dados, ou seja, o processo de obtenção do dado (gerado ou coletado); (b) Particularidades ou limitações sobre os dados que outros usuários devem conhecer; (c) Data da geração do conjunto de dados, condições de laboratório, quem preparou os dados, configurações de parâmetros, nome e versão do software utilizado; (d) Explicitar se são dados brutos ou processados; e (e) A versão dos dados arquivados e/ou reutilizados deve ser claramente especificada e documentada.

R1.3. (Meta)dados devem estar alinhados com padrões relevantes ao seu domínio	Atender aos padrões específicos da comunidade da área. Atender às boas práticas de arquivamento e ao compartilhamento do campo de pesquisa específico.
--	---

Fonte: (HENNING; RIBEIRO; SALES *et al.*, 2019).

De acordo com Imming (2018, p. 4), os princípios não são padrões, mas “[...] um guia de preparação dos dados de pesquisa para o reúso, sob condições claramente descritas por pessoas e máquinas”. Mons (2018, p. 3, *tradução nossa*), entretanto, confirma essa colocação, ao afirmar que o objetivo dos princípios é contextualizar e apontar para uma direção de maior utilidade e melhores serviços de dados, oferecendo suporte à sua reutilização e, assim, facilitando a escolha de quais padrões podem ser utilizados para esses fins.

Importa ressaltar que, apesar de inspirados no contexto da Ciência Aberta, os princípios FAIR não tratam de questões morais e éticas relativas à sua abertura, isto é, os dados não precisam, necessariamente, estar abertos. Mons (2018, p. 57, *tradução nossa*) explica que “[...] O ‘A’ do FAIR representa ser acessível sob premissas bem definidas, enquanto as condições de reutilização são cobertas pelo requisito de ter uma licença legível por máquina, de acordo com o princípio R1.1”.

No entanto, como afirma Henning *et al.* (2019), ainda que recentes no mundo científico nacional e internacional, quando colocados em prática tais princípios podem gerar equívocos devido às dificuldades no seu entendimento e na falta de experiência na sua aplicação e uso. Por esse motivo, conforme afirmam os autores, a iniciativa *GO FAIR*, apoiada pelos governos da Holanda, da Alemanha e da França, surge para ofertar um modelo de implantação desses princípios. Nesse sentido, em 2018, publica-se o Relatório Final e Plano de Ação intitulado *Turning FAIR into Reality*, que é um documento norteador para aplicação dos princípios FAIR em diferentes contextos (COLLINS *et al.*, 2018). Esse relatório discorre sobre o ecossistema dos dados FAIR, que são: políticas, planos de gestão de dados, identificadores persistentes, padrões e repositórios. Tais componentes figuram nesta pesquisa, conforme apresentado nos objetivos (Seção 1.3), com exceção dos planos de gestão de dados, os quais não são objetos deste estudo. Entende-se que semelhantes elementos, assim como apresenta o relatório, são essenciais para tornar os dados *FAIR* e *Open*, na medida que se inserem no contexto do compartilhamento dos dados.

Para Allen e Hartland (2018), a aplicação desses princípios é abstrata, uma vez que as áreas do conhecimento podem interpretar os princípios de forma diferente. No relatório apresentado pelos autores indicam-se as diferenças disciplinares como fator a ser observado

durante o uso dos princípios FAIR. Para tanto, o documento ressalta que aspectos políticos (orientação sobre gerenciamento de dados, conscientização e compreensão do FAIR), econômicos (implicações econômicas), social (propriedade dos dados, motivações e desafios) e técnicos (tipos de dados, armazenamento de dados, formatos e padrões, implementação) influenciam no uso do FAIR. Portanto, de acordo com Henning *et al.* (2019), a aplicação dos princípios é relativamente difícil, porque eles não foram devidamente detalhados e explicados na sua publicação original.

Quadro 7 - Os princípios FAIR

PRINCÍPIO FAIR	OS METADADOS DEVEM	DESCRIÇÃO
F1	Identificadores globais, persistentes e identificáveis.	Apresentar identificadores globais, persistentes e identificáveis.
F3	Incluir clara e explicitamente os identificadores dos dados que descrevem.	Explicitar o identificador único e persistente.
F4	Ser registrados ou indexados em recursos que ofereçam capacidades de busca.	Ser indexados em mecanismos de busca que possibilitem aos usuários encontrá-los com facilidade.
A1	Ser recuperáveis pelos seus identificadores usando protocolo de comunicação padronizado.	Protocolos de comunicação padronizados, i.e, HTTP ou Ftp.
A2	Ser acessíveis, mesmo quando os dados não estiverem mais disponíveis.	Os metadados devem ser sempre acessíveis.
I1	Ser representados por meio de uma linguagem formal, acessível, compartilhada e amplamente aplicável para a representação do conhecimento.	Adotar linguagens de representação do conhecimento que sejam padronizadas, acessíveis e amplamente aplicáveis. (p. ex.: RDF, XML, DICOM etc.).
I2	Usar vocabulários de acordo com os princípios FAIR.	Possuir referências a vocabulários e/ou ontologias que os descrevam e sigam os princípios FAIR.
I3	Incluir referências qualificadas para outros metadados.	Deve-se especificar se um conjunto de dados se baseia em outro.
R1	Ser descritos com uma pluralidade de atributos precisos e relevantes.	Informar quem pode ter acesso a eles, com que finalidade e sob quais condições. Especificar a proveniência dos dados. Atender aos padrões específicos da área.
R1.1	Ser liberados com licenças de uso de dados claras e acessíveis.	Definir explicitamente quem pode ter acesso a eles, com que finalidade e sob quais condições.
R1.2	Estar associados à proveniência detalhada.	De onde os dados vieram, qual a origem, foram reutilizados e/ou como o autor deseja ser reconhecido.
R1.3	Fazer parte de domínios com uso de padrões compartilhados em comunidades.	Dar atenção às boas práticas de arquivamento e compartilhamento específicos da área de pesquisa.

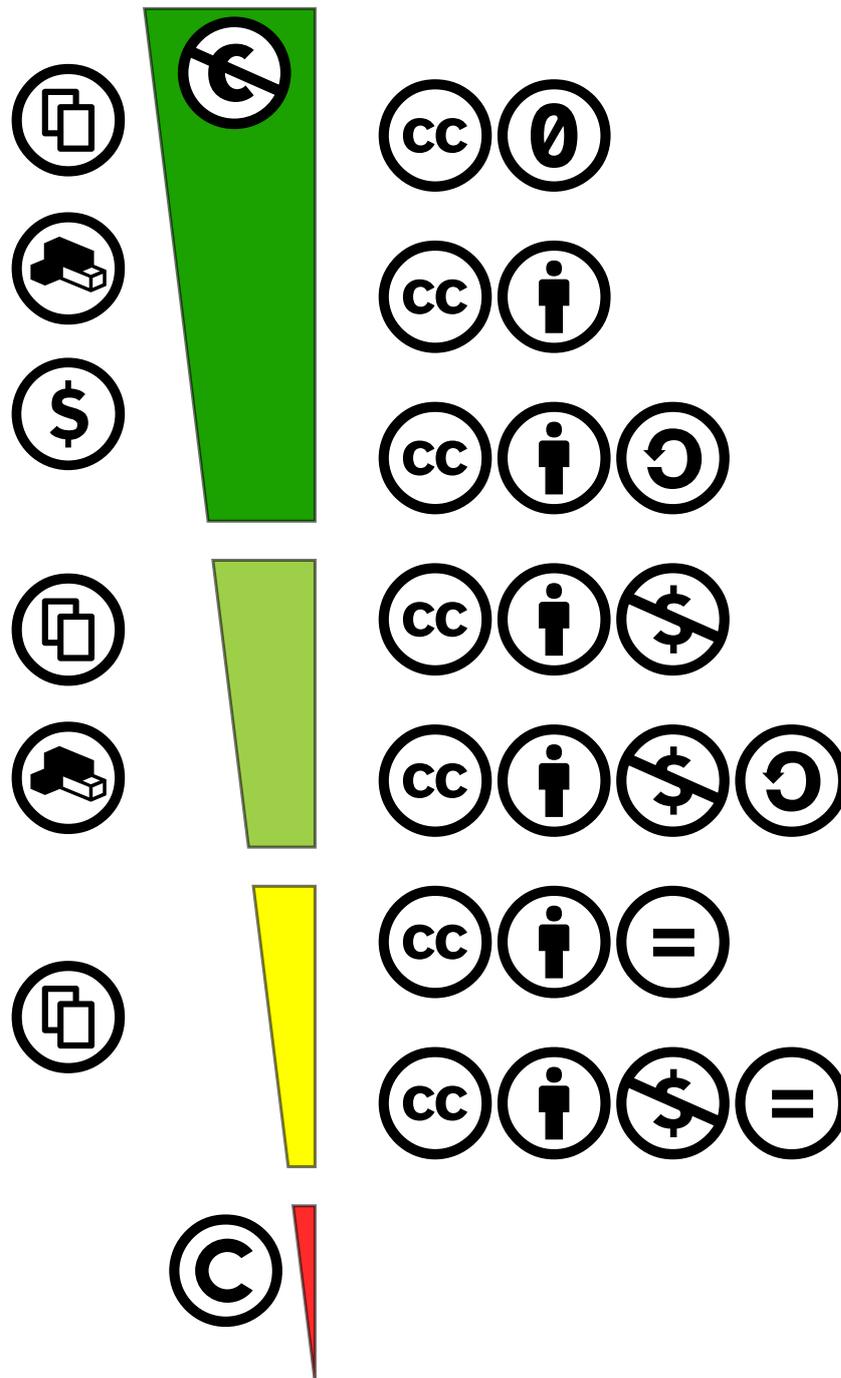
Fonte: Elaborado a partir das definições de Wilkison *et al.* (2016) e Veiga (2019).

Com relação ao princípio R1.1, sobre atribuição de licenças no contexto de Acesso Aberto, importa destacar que, apesar da legislação sobre direitos autorais variar de uma

jurisdição territorial para outra, eles exercem um papel central no Acesso Aberto, tendo em vista que a acessibilidade está sujeita à titularidade desses direitos (SWAN, 2016, p. 39).

A temática dos direitos autorais e licenciamento de dados é assunto extenso e complexo dentro do compartilhamento de dados. Não é objeto deste estudo aprofundar-se nesse tema, mas cabe ressaltar a importância e a existência das discussões sobre o contexto, uma vez que elas são apontadas pelos princípios FAIR como fatores para definir o nível *FAIRness* dos dados. Nesse sentido, Labastida e Margoni (2020) afirmam que, dentre as licenças para conteúdos abertos, as mais conhecidas são as licenças do *Creative Commons*. Apesar disso, segundo os autores, essas licenças não foram elaboradas especificamente para cobrir dados. Entretanto, a *Creative Commons*, entendendo tal demanda dos pesquisadores, educadores, artistas e outros criadores de conteúdo, idealizaram uma ferramenta legal chamada CC0, que renuncia a qualquer direito (CREATIVE COMMONS, 2020).

Figura 8 - Licenças CC em ordem de mais a menos abertas



Fonte: https://en.wikipedia.org/wiki/Creative_Commons_license.

Além das ferramentas fornecidas pelo *Creative Commons*, existem outros instrumentos jurídicos criados especialmente para dados, como os desenvolvidos pelo *Open Data Commons*, ou por alguns governos (LABASTIDA; MARGONI, 2020). *Open Data Commons* é uma solução jurídica para dados abertos, um projeto da OKF que tem o objetivo de beneficiar a comunidade de conhecimento aberto. A primeira licença de dados abertos foi lançada em março de 2008, tendo sido intitulada a *Public Domain Dedication and License*

(PDDL). A PDDL aplica aos dados e bancos de dados o domínio público, oferecendo um amplo grau de flexibilidade para os seus usuários (HATCHER, 2008). De fato, a atribuição de licenças aos dados de pesquisa se relaciona ao compartilhamento deles, uma vez que a presença da licença expõe aos outros usuários como o reuso dos dados pode ocorrer.

Diante de tal cenário, conforme afirma Henning *et al.* (2019), algumas iniciativas internacionais vêm elaborando métricas específicas que medem o grau de *FAIRness* dos objetos digitais. Os autores elucidam que o termo *FAIRness* significa o grau em que um recurso digital adere aos princípios FAIR. Essas métricas vêm sendo desenvolvidas de forma colaborativa para descrever as características, atributos e comportamentos (HENNING *et al.*, 2019). Contudo, Collins *et al.* (2018) afirma que, apesar de recursos importantes, as escalas de avaliação FAIR devem ser desenvolvidas com vistas a reconhecer e recompensar os diferentes níveis de FAIR, em vez de impor aos pesquisadores um *checklist* para que seja cumprido um nível mínimo de *FAIRness*. Os autores acrescentam que, sem uma discussão aprofundada e um nível de maturidade adequado, os pesquisadores irão se tornar relutantes em adotar os princípios FAIR. Eles ainda defendem que as comunidades científicas precisam apoiar as métricas FAIR para que elas possam ser aplicadas às práticas de compartilhamento de dados, respeitando as características de cada disciplina.

Em certa maneira, o nível *FAIRness* podem ser observados nos dados e metadados, porque, de acordo com os princípios FAIR, dados e metadados devem ser encontrados por pessoas e mecanismos automatizados. As limitações sobre o uso de dados, os protocolos para consulta e reuso de dados devem ser explicitados tanto para pessoas quanto para os mecanismos. Tais medidas facilitam e incentivam, em certa medida, o compartilhamento de dados.

2.2.4 *Compartilhamento de dados de pesquisa*

Diante do desenvolvimento da web, das grandes redes e sistemas de informação, percebem-se transformações nas atividades do ser humano. Para os autores Medeiros e Caregnato (2012, p. 312), uma dessas mudanças “ocorre devido à possibilidade de colaboração proporcionada por redes distribuídas, favorecendo que pesquisadores possam, entre outros aspectos, realizar estudos de forma a compartilhar e aproveitar dados de pesquisas anteriores”.

Assim, o compartilhamento de dados pode ser entendido como uma forma de colaboração. A colaboração pode ser entendida, em linhas gerais, como “trabalho feito em

comum com uma ou mais pessoas; cooperação, ajuda, auxílio [...] trabalho, ideia, doação etc. que contribui para a realização de algo ou para ajudar alguém; auxílio [...] participação numa obra literária, científica etc.” (HOUAISS, 2019). No contexto da ciência, Sonnenwald (2008, p. 645) apresenta o conceito de colaboração como um tipo de “interação que ocorre dentro em um contexto social entre dois ou mais cientistas, facilitando o compartilhamento de significado e de conclusão das tarefas em relação a um objetivo compartilhado mutuamente”. Zanetti (2012) define compartilhar como “participar de algo”, “tomar parte em alguma coisa”, partilhar, dividir com outros etc.

As novas práticas de fazer ciência são observadas por Gray (2009), o qual argumenta que a ciência nasce empírica e evolui para uma ciência capaz de criar e utilizar modelos para gerar teorias sobre os fatos do mundo, com vistas a simular fenômenos complexos e explorar dados por meio de computadores. Complementarmente, constata-se que o modo de fazer pesquisa, ou seja, a metodologia, também “está mudando, e estamos no limiar de uma nova era da ciência dirigida por dados” (HEY; HEY, 2006).

Para Zanetti (2012), o conceito de compartilhamento ganha estímulo com o advento da Internet, à medida que ela foi se tornando mais organizada e sistematizada. Ainda segundo a autora, a sistematização ocorreu não apenas por causa do aprimoramento do aparato tecnológico, mas essencialmente devido a um discurso legitimador e ao mesmo tempo aglutinador em torno da prática do compartilhamento. Em paralelo a esse discurso, convém ressaltar que o surgimento da Internet trouxe a ideia de “democratização” da informação por meio de sua descentralização, isto é, permitiu que a informação possa ser produzida e acessada por muitos (LEMOS, 2002; LÉVY, 2010).

No contexto de colaboração científica, destaca-se o tema da *e-Science*, definida como estrutura que objetiva a colaboração entre cientistas a partir do compartilhamento e gerenciamento de dados científicos primários (MEDEIROS; CAREGNATO, 2012). Os autores destacam que a *e-Science* altera fundamentalmente a maneira como 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. Nesse sentido, Medeiros e Caregnato (2012) argumentam que a possibilidade de tratar dados, em vez de coletá-los novamente, permite que os esforços sejam focados na análise dos dados já existentes e compartilhados a fim reduzir custos e esforços, e possibilitar avanços mais rápidos. Portanto, os autores concluem que a *e-Science* é orientada ao compartilhamento e reúso de dados.

A discussão sobre o acesso aos dados de pesquisa se faz presente na comunidade científica em razão dos avanços tecnológicos e da nova tendência de compartilhar e produzir conhecimento científico. Outras comunidades, como sociedades profissionais, financiadores, formuladores de políticas e editores, ressaltam a importância e os benefícios do compartilhamento de dados de pesquisa (DRAZEN *et al.*, 2016; NATIONAL INSTITUTE OF HEALTH, 2015). Além do mais, segundo Sales (2014, p. 49), pesquisadores, editores científicos, agências de fomento à pesquisa e instituições acadêmicas reconhecem que os dados, quando preservados e gerenciados, são fonte de recurso informacional.

Muito se fala sobre os benefícios do compartilhamento de dados, e isso motivou muitas organizações – as quais apoiam a pesquisa – a exigir que os dados sejam disponibilizados publicamente (VASILEVSKY *et al.*, 2017). Por definição do *Open Knowledge Foundation*, o compartilhamento de dados de pesquisa permite aos pesquisadores replicar, verificar e expandir estudos concluídos sem nenhuma barreira, além de ajudar a impedir e detectar fraudes relacionadas aos dados.

Desde 2011, a NSF passou a exigir dos seus candidatos o envio de um plano de gerenciamento de dados. O objetivo é que os pesquisadores documentem quais dados primários e recursos de pesquisa serão compartilhados com outros pesquisadores (NATIONAL SCIENCE FOUNDATION, 2007). Ao redor mundo, observa-se um movimento em prol de institucionalizar o compartilhamento de dados de pesquisa. Por exemplo, o Escritório de Política Científica e Tecnologia da Casa Branca (*White House Office of Science and Technology Policy*) emitiu um memorando em 2013 para orientar as agências de fomento a elaborar planos de gerenciamento de dados a fim de garantir o acesso público aos resultados de pesquisa financiados pelo governo federal, incluindo dados (ESTADOS UNIDOS DAS AMÉRICAS, 2013). Nesse sentido, o estudo de Faniel e Jacobsen (2010, p. 355) aponta que, desde 2010, em países como Estados Unidos e Inglaterra, por exemplo, “o compartilhamento e gerenciamento de dados têm sido o foco principal entre as agências federais de financiamento, universidades e comunidades científicas”.

Em 2014, o NIH implementou uma política de compartilhamento de dados para dados genômicos humanos e não humanos em larga escala (VASILEVSKY *et al.*, 2017). Além disso, o Conselho Europeu de Pesquisa incluiu nas Diretrizes de Acesso Aberto o acesso do público aos dados de pesquisa (MALAPELA, 2016).

O compartilhamento de dados de pesquisa é um tema que se consolida dentro de todas as esferas que envolvem a pesquisa científica, não apenas no campo do acesso aberto.

O compartilhamento de dados para reutilização ou reprodução de experimentos exige conhecer sua origem e entender como foram produzidos, associando à informação métodos, algoritmos ou técnicas adotados, e ainda ter acesso ao software necessário para processá-los, o que torna o processo bastante complexo. Sem isso, pode não ser possível reproduzir o experimento original ou reutilizar o dado em uma outra pesquisa (MARQUES, 2014, p. 56).

O compartilhamento e reuso de dados permite economia no tempo da pesquisa, pois os dados já estão coletados (MEDEIROS; CAREGNATO, 2012). De acordo com os autores, isto contribui também para a produção do conhecimento científico – pois não há necessidade de retroceder à parte que já foi executada de um processo –, bem como evita que recursos já concedidos por agências de fomento sejam duplicados. Dentre as vantagens do compartilhamento de dados, destacam-se as propostas por Giglia e Swan (2012), conforme o Quadro 8:

Quadro 8 - Vantagens do compartilhamento dos dados abertos para a ciência e para a sociedade

PARA A CIÊNCIA É POSSÍVEL:	PARA A SOCIEDADE É POSSÍVEL:
Acessar os dados para compará-los, contrastá-los e debatê-los. Podemos ter uma ciência mais sólida baseada em competitividade.	Reutilização para inovação e aumento da abertura dos dados pode ajudar a gerar novos produtos.
Adicionar novos dados para criar novo conhecimento. Os dados abertos são mais visíveis, se propagam mais rapidamente, aceleram a criação do conhecimento e promovem o progresso da ciência.	Reutilização para a criação de valores por intermédio de mineração de dados. A exploração desse vasto recurso de dados e informação pode gerar benefícios para a economia, aprimorando a produtividade e o valor adicional para o consumidor.
Reutilizar os dados para novas pesquisas sem custos para coletá-los novamente. Uma vez coletados, analisados e depositados, poderão ser usados em novas pesquisas em diferentes perspectivas.	Reutilização para as novas mixagens. A abertura dos dados pode ser utilizada tanto em aplicações para computador e dispositivos móveis.
Reutilizar os dados para verificar e detectar a fraude e/ou a falsificação. Mais transparência poderá evitar fraude.	Reutilização para preservação. A preservação será mais fácil se os dados estiverem abertos e interoperáveis via padrões comuns.
Criar novos dados com técnicas de mineração de dados. Ao ser aplica a dados abertos e licenciados para reutilização proporciona aumento de valor e economia de tempo.	Reuso para formulação de tomada de decisão. Os formuladores de políticas podem ser capazes de tomar decisões baseadas em evidências e monitorar o impacto dessas decisões.
	Reutilização para o conhecimento da sociedade. A abertura de dados permite que os cidadãos sejam mais informados nos níveis político, cultural e científico.

Fonte: (HENNING *et al.*, 2019).

No entanto, o compartilhamento de dados e a administração a longo prazo envolvem uma série de atividades, participantes e tecnologias, especialmente se a descoberta, a reutilização e a preservação forem asseguradas (DALLMEIER-TIESSEN *et al.*, 2014). Nesse sentido, uma das discussões em relação ao compartilhamento de dados é o fato de que o ônus recai sobre o pesquisador, que precisa tratar seus dados, além de seguir convenções e normas de publicações, de acordo com sua área de conhecimento. Em contrapartida a esse ônus, os estudos desenvolvidos por Piwowar, Day, Fridsma (2007), Mueller-Langer e Andreoli-Versbach (2018) revelam que há um aumento do número de citações para aqueles que publicam os dados.

Um levantamento realizado em 2010, com 1.300 cientistas ao redor do mundo, mostrou que mais de 80% dos entrevistados gostariam de utilizar conjuntos de dados de outros cientistas. No entanto, em 2011, somente 36% afirmaram que disponibilizam os próprios dados de maneira amigável (TENOPIR *et al.*, 2011). O estudo realizado pelos autores investigou as práticas atuais de compartilhamento de dados e as percepções das barreiras e facilitadores do compartilhamento. Estima-se que o alcance da entrevista foi de 15 mil pesquisadores (no caso, a taxa de resposta é de aproximadamente 9%). A maioria dos entrevistados são acadêmicos que dividem seu tempo entre pesquisa, ensino, administração e outras atividades, e representam uma variedade de disciplinas de ciências sociais, de modo que a maioria das disciplinas-alvo pertencem às ciências ambientais e à ecologia.

A publicação e/ou o compartilhamento de dados pelos pesquisadores diverge entre as disciplinas acadêmicas. Uma pesquisa da editora científica Wiley investigou a probabilidade de os pesquisadores compartilharem dados, de modo que foi enviado um questionário para 90 mil pesquisadores ao redor do mundo. Eles obtiveram mais de 2.250 respostas. De acordo com o estudo, 66% dos *cientistas da vida*⁸ compartilham seus dados com outros pesquisadores, principalmente em repositórios de dados ou em periódicos, enquanto apenas 36% dos cientistas sociais e estudiosos de humanidades compartilham seus dados. Porém, muitas vezes tais dados são compartilhados informalmente, ou seja, dentro de sua própria equipe de pesquisa ou instituição (FERGUSON, 2014; MEADOWS, 2014).

De acordo com Andreoli-Versbach e Mueller-Langer (2014), em estudo realizado na área de economia e administração, existem incentivos das instituições, das agências de fomentos e dos periódicos científicos para compartilhar dados de pesquisa. Contudo, os

⁸ De acordo com o estudo, são os pesquisadores da área de medicina, biomedicina, biologia, genética etc.

pesquisadores não disponibilizam os dados de forma significativa, mas admitem que compartilhar os dados colabora para o progresso científico.

O cenário apontado pelos autores também se dá nas outras áreas do conhecimento, como nos estudos de LeClere (2010), na área das ciências sociais, e de Longo e Drazen (2016), na área da medicina, bem como na pesquisa desenvolvida por Savage e Vickers (2009) com autores que publicaram no periódico *The Public Library of Science (PLoS) Journals*. Em resumo, os estudos relatados pelos autores indicam alguns dos obstáculos para o compartilhamento de dados, como a falta de financiamento, o medo de críticas ou má interpretação, a concorrência, a dificuldade de navegar nas ferramentas de infraestruturas e a escassez de recompensas pelo compartilhamento de dados.

Para Meadows (2014), é possível delinear os três desafios a serem vencidos para o compartilhamento de dados, baseados na pesquisa realizada pela Wiley. Primeiro, os editores ao longo dos anos vêm hospedando dados na forma de documento suplementar, de maneira arbitrária e não planejada. No entanto, segundo a autora, essa prática não é adequada para curadoria e preservação a longo prazo dos dados da pesquisa. Sem dúvida, pode-se fazer mais – por exemplo, exigindo que os autores arquivem seus dados em repositórios de dados. De acordo com Meadows (2014), tal prática já é adotada por periódicos e comunidades científicas, como o periódico *Molecular Ecology*, que introduziu uma política de arquivamento de dados em 2011; e as comunidades científicas *American Geophysical Union*, a *Society for the Study of Evolution* e a *British Ecological Society*, que introduziram políticas de compartilhamento de dados para seus periódicos. Logo, a maioria dos periódicos sobre genética exige que os autores depositem quaisquer sequências de DNA mencionadas no GenBank.

O segundo desafio apresentado por Meadows (2014) é a falta de requisitos da agência de fomento à pesquisa. Conforme a autora, os governos e outros órgãos de financiamento precisam desempenhar um papel mais ativo. Embora alguns já exijam o envio de planos de gestão de dados, estes nem sempre são aplicados, e muitos órgãos ainda não elaboraram os requisitos. Outro ponto é o apoio e o investimento aos repositórios de dados. De acordo com a pesquisadora, os repositórios gerais e específicos da disciplina oferecem aos pesquisadores uma maneira simples, segura e permanente de compartilhar seus dados, mas eles estão subutilizados.

O último desafio apresentado por Meadows (2014) é encontrar maneiras de creditar os pesquisadores pelo compartilhamento de dados – novamente, uma área em que instituições

e financiadores precisam mostrar liderança. Esse último obstáculo exigirá não apenas a adoção de padrões de citação de dados, mas também uma mudança cultural na forma como os pesquisadores são recompensados por criar, analisar e preservar dados. Encontrar tais maneiras é um passo importante no processo de incentivo ao compartilhamento de dados. Nesse sentido, Brase, Lautenschlager e Sens (2015), ao se referirem ao uso de identificadores persistentes para dados de pesquisa, também afirmam que os cientistas estão preparados para fornecer dados, mas não há reconhecimento pelo trabalho de processar, produzir documento de apoio, glossário de dados e garantir a qualidade dos dados.

Várias justificativas sustentam os argumentos para melhor acesso e curadoria de dados de pesquisa (BORGMAN, 2012). Ao disponibilizar dados primários, outros têm a oportunidade de replicar ou corrigir descobertas anteriores e, ostensivamente, influenciar o ritmo e a eficiência de futuros empreendimentos de pesquisa (VASILEVSKY *et al.*, 2017). Os pesquisadores podem fazer novas perguntas sobre os dados existentes, e os dados podem ser combinados e organizados com vistas a aumentar seu valor (BORGMAN, 2012). Como Fischer e Zigmond (2010) argumentam, os grandes avanços da ciência dependem não somente das contribuições dos pesquisadores individuais, mas também da disposição em compartilhar os produtos do trabalho.

Os pesquisadores Torres-Salinas, Robinson-Garcia e Cabezas-Clavijo (2012) relatam quais as vantagens do compartilhamento de dados. Dentre elas, destaca-se o aproveitamento de recursos financeiros investidos em ciência, que permite o reuso de dados para novas análises; a identificação de fraudes, que permite replicar experimentos e verificar as hipóteses apresentadas em determinadas áreas do conhecimento; e o benefício pessoal para os pesquisadores, uma vez que os dados podem ser citados e, com isso, o número de citações aumenta.

Com o passar do tempo, o desenvolvimento de políticas de compartilhamento de dados de pesquisa se torna cada vez mais popular, pois, conforme afirmam Piwowar e Chapman (2008), para alcançar uma cultura de compartilhamento de dados são necessárias diretivas e normas. Em vista disso, as agências de fomento e periódicos começam a solicitar e, em alguns casos, exigir que os pesquisadores compartilhem seus conjuntos de dados com outros pesquisadores.

Segundo Niu (2009), uma das maneiras de mensurar os benefícios do compartilhamento de dados se dá pelo seu reuso. De acordo com Medeiros e Caregnato (2012), “o reuso de dados científicos pode ser visto, a grosso modo, como a utilização de

dados já utilizados, ou seja, são dados utilizados mais de uma vez, mas com intenções originais”.

De acordo com Faniel e Jacobsen (2010), em geral, há três perguntas que cientistas consideram ao avaliar a reusabilidade de dados: (i) os dados são relevantes?; (ii) os dados podem ser entendidos?; e (iii) os dados são confiáveis?

Quadro 9 - Reusabilidade dos dados de pesquisa

Perguntas	
os dados são relevantes?	refere-se à avaliação que os cientistas fazem sobre o nível em que os dados estão para atender seus problemas de pesquisa
os dados podem ser entendidos?	diz respeito à possibilidade de entender os dados, em geral, a partir de anotações de metadados e documentação disponível
os dados são confiáveis?	diz respeito à possibilidade de entender os dados, em geral, a partir de anotações de metadados e documentação disponível

Fonte: Adaptado de Faniel e Jacobsen (2010).

Faniel e Jacobsen (2010), por sua vez, descrevem que a maior dificuldade encontrada para o reuso de dados é identificar o contexto de produção dos mesmos. Segundo os autores, as dificuldades para reuso de outros cientistas são conhecer as condições em que os dados foram produzidos, a interlocução entre quem produz e quem necessita de dados, e a crença de que o seu fornecimento amplo por si só irá garantir a reutilização. Nesse sentido, a proposta FAIR se preocupa com tal aspecto, como elucidado na subseção anterior, pois, ao tornar os dados FAIR, pressupõe-se que essa informação esteja disponível.

De acordo com Gray (2009, p. xvii), parte importante do processo de compartilhamento e reuso é a ideia de que, após os “dados serem capturados, necessita-se que sejam curados antes que se inicie qualquer tipo de análise”. Dessa forma, como ressaltam Soehner, Steeves e Ward (2010), os centros devem estabelecer estratégias para curadoria dos dados, isto é, definir como tratar os dados desde a sua conceitualização ou recebimento até o uso e a preservação. Nesse sentido, entende-se que o compartilhamento de dados está ligado à curadoria deles, uma vez que a curadoria de dados é o processo de gerenciamento de dados científicos primários que permite, a partir da preservação e manutenção, agregar valor aos dados de pesquisa e prover o compartilhamento e o reuso (MEDEIROS; CAREGNATO, 2012).

1.1.1.1 Curadoria de dados de pesquisa

De acordo com Sayão e Sales (2016b), “o ponto crucial da curadoria de dados de pesquisa é que ela precisa estar inserida organicamente nos fluxos de geração de conhecimento científico para ter validade”. Os autores esclarecem que os profissionais de

informação e as infraestruturas informacionais e tecnológicas devem estar imbricadas nas atividades dos laboratórios, das outras atividades acadêmicas e de pesquisas da instituição.

A curadoria se inicia no planejamento dos dados e não se encerra com o fim dos projetos, pois os dados continuam a evoluir (SAYÃO; SALES, 2016b). Dessa forma, é necessário que o profissional da informação se preocupe com a gênese dos dados e com toda a documentação necessária à interpretação e à contextualização dos dados ao longo do tempo (SAYÃO; SALES, 2016b). Os autores defendem, ainda, que a curadoria de dados de pesquisa exige uma equipe com muitas expertises que vão além da biblioteconomia de dados.

Os diferentes tipos de dados requerem processos distintos. Por exemplo, como afirmam Sayão e Sales (2016b), os dados observacionais precisam ser preservados e arquivados de modo que propriedades arquivísticas, como proveniência, confiabilidade e autenticidade sejam mantidas, colocando em destaque o profissional chamado arquivista de dados. Como complemento, os autores apresentam os cientistas de dados, que no contexto da curadoria são os cientistas da computação. Esses profissionais precisam analisar e processar dados por meio de programas e equipamentos específicos que identificam padrões auxiliares na formulação de hipóteses.

Portanto, a curadoria implica, em grande escala, em:

[...] adicionar valor aos dados. Nessa direção, os dados precisam ser avaliados, analisados, enriquecidos com anotações, ligados por hiperlinks com outros recursos, comentados, agregados, “regerados” e descartados. Este papel, que exige conhecimento profundo da área, é atribuído ao pesquisador e a outros especialistas da área (SAYÃO; SALES, 2016b).

De acordo com Digital Curation Centre (DCC) (2019, *tradução nossa*), curadoria digital de dados é a “manutenção, preservação e atribuição de valor a dados digitais de pesquisa durante todo o seu ciclo de vida. Este gerenciamento efetivo diminui os riscos de obsolescência e desvalorização da pesquisa ao longo do tempo”. Afirma-se, também, que a curadoria digital está relacionada com o modo de tornar dados disponíveis em formato digital para acesso futuro. A ideia é tornar esses dados acessíveis, interpretáveis, reutilizáveis e confiáveis (NEUROTH *et al.*, 2013). Tais afirmações alinham-se aos princípios FAIR, elucidados na subseção anterior, na medida em que tornar os dados FAIR significa que os dados podem ser encontrados, estão acessíveis, interoperáveis e recuperáveis (*Findable, Accessible, Interoperable, and Reusable*).

Lee e Tibbo (2007, *tradução nossa*) afirmam que a curadoria digital é responsável por:

Impulsionar a disponibilidade de dados para a comunidade científica; proporcionar armazenamento redundante; transformar os dados; preservar a longo prazo; disponibilizar dados digitais autênticos para serem reproduzidos e reutilizados; desenvolver repositórios digitais confiáveis e duráveis; princípios de criação e captura de metadados; uso de padrões abertos para formatos e conversão de arquivos.

Já de acordo com Sayão e Sales (2012), curadoria digital é um conceito cunhado a partir do conjunto de estratégias, abordagens tecnológicas e atividades resultantes do acúmulo de conhecimentos, práticas em preservação e acesso a recursos digitais realizadas na última década. Os autores afirmam que é um conceito em evolução. Contudo, o termo “curadoria digital”, segundo os autores, já tem como ideia central de gestão atuante a preservação de recursos digitais durante todo o ciclo de vida de interesse do mundo acadêmico e científico.

Por seu lado, Abbott (2008) expande o conceito de curadoria digital, definindo-a como atividades envolvidas na gestão de dados, desde a criação, passando pelas boas práticas na digitação, na seleção dos formatos e na documentação, além da disponibilização e da garantia de reuso no futuro (SAYÃO; SALES, 2012).

A curadoria digital também inclui a gestão de grandes conjuntos de dados para uso diário, assegurando, por exemplo, que eles possam ser pesquisados e continuem viáveis, ou seja, capazes de ser lidos e interpretados continuamente. Nessa perspectiva, a ideia de curadoria digital estende-se além do controle do repositório que arquiva os recursos e envolve a atenção do criador do conteúdo e dos usuários futuros. Corrêa e Bertocchi (2012, p. 29) destacam a evolução e contextualização da prática da curadoria digital, pois, conforme a expansão da sociedade digitalizada, o termo “curadoria” passou a ser utilizado amplamente para as atividades que envolvem a gestão de dados. Portanto, compreende-se que semelhante tema é amplo, extrapolando os objetivos desta pesquisa, que é apenas o contexto do compartilhamento de dados vinculados a periódicos científicos. Contudo, as questões de curadoria de dados levaram aos avanços nas discussões acerca do compartilhamento e comunicação dos dados, exatamente por ser um conceito que perpassa essas temáticas (ABBOTT, 2008; CORRÊA; BERTOCCHI, 2012; SAYÃO, SALES, 2012). Diante desse contexto, observam-se iniciativas para promover o acesso aos dados de pesquisa, entendendo a importância de tal processo para a ciência.

1.1.1.2 Iniciativas para a promoção do acesso aos dados de pesquisa

Em 2004, OECD apresentou um documento intitulado *Declaration on Access to Research Data from Public Funding*. A declaração foi assinada por 34 países, os quais devem seguir as dez diretrizes (Quadro 10), e reitera a importância da comunicação dos dados de pesquisa para o desenvolvimento da ciência. Três anos depois, em 2007, a OECD publicou o *OECD Principles and Guidelines for Access to Research Data from Public Funding*, cujo objetivo é apresentar as ações coordenadas a nível internacional, necessárias para disseminação do acesso aberto a dados de pesquisa financiada por órgãos públicos, contribuindo para o avanço da ciência e da inovação (OECD, 2007). Neste último documento acrescentam-se quatro novos princípios aos dez criados em 2004 (Quadro 10).

Quadro 10 - Princípios propostos pela OECD para abertura dos dados de pesquisa financiadas

ANO	PRINCÍPIO	DEFINIÇÃO
2004	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 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 de 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.

ANO	PRINCÍPIO	DEFINIÇÃO
	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.
2007	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: Costa (2017, p. 58-59).

Em 2009, o grupo de trabalho *Work Grouping Open Data in Science (WGODS)* – criado a partir do OKF e tendo como objetivo facilitar o acesso e uso de dados em vários contextos – declarou ter por missão orientar e criar ferramentas que possibilitem o compartilhamento e reúso dos dados de pesquisa (COSTA, 2017). Segundo a coordenadora do WGods, Jennifer C. Molloy, embora existam problemas culturais e inércia em face da Ciência Aberta, observa-se que as agências de fomento estão requisitando planos de gestão de dados como condição de financiamento de pesquisa (MOLLOY, 2011). Para Costa (2017,

p. 59), a criação do grupo de trabalho marcou a ampliação da noção da abertura de dados em relação à filosofia aberta, visto que esse grupo sistematizou um conjunto de condições que define a abertura do acesso e do uso de dados.

A discussão proposta pelo grupo de trabalho amplia a noção de Ciência Aberta, problematizando as licenças atribuídas aos recursos abertos, que devem ser mais do que gratuitos. Após dois anos, o grupo de trabalho desenvolveu um conjunto de princípios para publicação dados científicos abertos, usando o conceito de open definido pela OKF e o Science Commons' Protocol for Implementing Open Access Data (Protocolo da Science Commons para implementação de dados abertos) como modelo e lente teórica. O resultado foram os Princípios Panton, cujo escopo cobre todos os dados experimentais primários publicados dentro ou em paralelo aos artigos científicos, incluindo o conteúdo de dados de qualquer tabela ou gráfico; e todas as imagens, áudio ou vídeo, atuando como o mecanismo principal de captura de dados, por exemplo, géis de proteína ou gravações de animais (MOLLOY, 2011). Para a autora, a mensagem principal dos Panton é que todos esses dados - com muitas poucas exceções - devem ser colocados explicitamente no domínio público.

A Organização Mundial da Saúde (OMS) publicou o documento *Welcome Trust*, que se consolida como um código de conduta e descreve os requisitos para gerenciamento e acesso a dados criados por pesquisas financiadas. Nesse mesmo ano, os *Research Councils United Kingdom* (RUCK) publicaram um novo conjunto de princípios para a comunicação de dados de pesquisa, intitulado *RUCK Common Principles on Data Policy*. O documento, conforme esclarecem Corti *et al.* (2014), representa o avanço na promoção do compartilhamento de dados de pesquisa, pois a RUCK é responsável por mais da metade dos financiamentos de pesquisas da região.

No contexto da União Europeia (*European Union*) destaca-se o programa Horizon 2020, que promove a comunicação de dados de pesquisa (COSTA, 2017). O foco é otimizar os investimentos da União Europeia em pesquisa e inovação. O documento, elaborado pelo programa, estabelece o direito de acesso e reuso dos dados de pesquisa, os deveres dos beneficiários de recursos da EU e a necessidade de elaboração de um plano de gestão de dados como cumprimento dos requisitos.

No mesmo ano, o governo dos Estados Unidos oficializa um memorando aos chefes dos departamentos e agências para aumentar o acesso aos resultados de pesquisas financiadas. Diante desse cenário, a NSF instaurou a obrigatoriedade de um Plano de Gerenciamento de Dados para as submissões de propostas de financiamento de pesquisa. Já

governo da Austrália criou o programa *Australian National Data Service* (AND) com vistas a desenvolver infraestrutura para dados de pesquisa e a fim de tornar os dados mais valiosos para pesquisadores, instituições de pesquisa e nação. Por meio desse programa foi criada a iniciativa *Australian Research Data Commons* (ARDC), que permite à comunidade de pesquisa e ao setor de pesquisa da Austrália acesso à infraestrutura, plataformas e conjuntos de dados de relevância nacional.

Retomando os Princípio Panton, importa destacar que não são uma iniciativa isolada, como afirma Molloy (2011), faz parte de um movimento mais amplo para promover dados abertos na ciência que está ganhando impulso. Para a autora os Princípios Panton podem ser resumidos em quatro tópicos: 1) Ao publicar dados, faça uma declaração explícita e robusta de seus desejos; 2) Use uma renúncia de direitos autorais reconhecida ou licença que seja apropriada para os dados; 3) Se o intuito é que os dados sejam efetivamente usados por outros, eles devem ser abertos conforme a definição proposta pela *Open Knowledge / Data Definition*

- em particular, cláusulas não comerciais e outras cláusulas restritivas não devem ser usadas; 4) Recomenda-se o uso de licenças explícitas aos dados científicos publicados para o domínio público via PDDL⁹ ou CCZero¹⁰, de maneira a garantir a conformidade com o *Science Commons' Protocol for Implementing Open Access Data* e com a *Open Knowledge / Data Definition*.

Nesse contexto de iniciativas para comunicação de dados, o estudo realizado por Costa (2017) apresenta um conjunto de princípios e recomendações para tal. Nessa pesquisa, a autora avaliou cinco documentos que tratam dos princípios e recomendações para a comunicação dos dados de pesquisa. Dos cinco documentos citados pela autora, quatro foram abordados nesta seção: OCD, WGOP, RUCK e H2020. Na análise de Costa (2017), os documentos apresentam 14 princípios relevantes para a comunicação dos dados de pesquisa, que buscam delinear um conjunto de princípios e recomendações a nível internacional (Quadro 11).

⁹ Disponível em: <http://opendatacommons.org/licenses/pddl/1-0/>

¹⁰ Disponível em: <http://creativecommons.org/publicdomain/zero/1.0/>

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 dos dados 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: Costa (2017, p. 72).

Na literatura referente ao tema do compartilhamento e da comunicação de dados de pesquisa há diversos estudos que investigam os fatores influenciadores de suas práticas, ressaltando que muitos desses estudos foram discutidos nesta subseção. Nesse sentido, destaca-se o estudo realizado por Kim e Stanton (2016), que investigou como os ambientes institucionais e as motivações individuais influenciam os comportamentos de

compartilhamento de dados dos cientistas em diversas disciplinas. Em seu estudo, constataram que as pressões institucionais (ou seja, a pressão regulatória por periódicos e a pressão normativa nas disciplinas) e as motivações individuais (isto é, o benefício percebido na carreira, o esforço percebido e o altruísmo acadêmico) têm relações significativas com os dados dos cientistas.

O trabalho realizado pelos autores elenca oito fatores que influenciam o compartilhamento de dados em diferentes áreas do conhecimento. Durante a discussão, a análise apresenta os estudos já realizados acerca do tema, ressaltando que algumas das pesquisas anteriores tinham o escopo de investigar apenas uma área do conhecimento específica (Quadro 12).

Quadro 12 - Os oito fatores que influenciam o compartilhamento de dados de pesquisa em diferentes áreas do conhecimento

	FATOR	DESCRIÇÃO
Nível de indivíduo	Benefício para carreira	Exerce uma influência significativa nos comportamentos de compartilhamento de dados dos cientistas. Isto é, os cientistas que percebem que há mais benefícios profissionais no compartilhamento de dados em seus artigos publicados têm maior probabilidade de compartilhar seus dados com outras pessoas.
	Riscos para carreira	Não se constatou uma relação significativa com compartilhamento de dados. Pois a pesquisa conceituou compartilhamento de dados apenas para dados de artigos publicados. Portanto, os diferentes conceitos de compartilhamento de dados precisam ser considerados durante a interpretação dos dados.
	Esforço	Teve um resultado negativo significativo sobre os comportamentos de compartilhamento de dados dos cientistas. Isto é, os cientistas que percebem que compartilhar dados exige um trabalho adicional são menos propensos a compartilhar.
	Altruísmo acadêmico	Tem uma relação significativa com os comportamentos de compartilhamento de dados dos cientistas. Isto é, o altruísmo do pesquisador exerce influência significativa no compartilhamento de informações.
Nível institucional	Políticas de compartilhamento de dados de agências de fomento	A pressão regulatória das agências financiadoras não demonstrou ter um relacionamento significativo com os comportamentos de compartilhamento de dados dos cientistas. Assim sendo, difere dos resultados de pesquisas anteriores, que argumentaram sobre o fato de que as políticas de compartilhamento de dados das agências financiadoras influenciam positivamente o compartilhamento de dados dos cientistas.
	Políticas de compartilhamento de dados de periódicos	Tem influência significativa nos comportamentos de compartilhamento de dados dos cientistas. Isto é, demonstra que os periódicos que adotam políticas de compartilhamento de dados exercem pressão sobre os pesquisadores para compartilhar.

	FATOR	DESCRIÇÃO
	Pressão normativa	A pressão normativa de cada disciplina científica (ou comunidade) influencia significativamente os comportamentos de compartilhamento de dados dos cientistas em diferentes disciplinas. Há variações significativas entre as disciplinas, porém os padrões e normas adotados influenciam positivamente o compartilhamento de dados dos cientistas.
	Repositório de dados	A disponibilidade de repositórios de dados em uma disciplina não foi encontrada como tendo um relacionamento significativo com os comportamentos de compartilhamento de dados dos cientistas.

Fonte: Adaptado de de Kim e Stanton (2016, *tradução nossa*).

Por fim, o compartilhamento de dados pode ser entendido, segundo Borgman (2012), como a ação realizada em prol da disponibilização de dados de pesquisa para o uso de outros. Com relação às formas de disponibilizar os dados, a autora assinala que entre as diferentes de maneiras há o uso de repositórios de dados.

2.2.5 *Repositório de dados*

Os repositórios digitais são “bases de dados on-line que [...] armazenam arquivos de diversos formatos” (IBICT 2019). Segundo Leite (2009, p. 19), repositórios digitais, no contexto do acesso aberto, são “provedores de dados que são destinados ao gerenciamento de informação científica, constituindo-se, necessariamente, em vias alternativas de comunicação científica”. Isto resulta em uma série de benefícios tanto para os pesquisadores quanto para as instituições ou sociedades científicas, o que proporciona maior visibilidade aos resultados de pesquisas e possibilita a preservação da memória científica de sua instituição.

A literatura apresenta uma diversidade de tipos de repositórios digitais. A classificação mais comum considera o objetivo para o qual o repositório é construído, dividindo-os em: i) Repositórios Institucionais, voltados para a o armazenamento da produção científica de uma determinada instituição; e ii) Repositórios Temáticos, voltados para a reunião da produção técnico-científica sobre uma determinada temática. Os repositórios institucionais estão inseridos no centro do movimento do acesso aberto à informação científica, pois são ferramentas que modificam o sistema de comunicação e o modo de gerenciar a informação (LEITE, 2009).

Sales (2014) afirma que o termo repositório, no âmbito da comunicação científica, apresenta o conceito com características que o diferencia de uma base de dados comum. Tais características permitem que o repositório cumpra a sua função de preservar e disseminar a

informação. A criação de repositórios de dados é um dos caminhos para a preservação de dados gerados pela pesquisa científica (CAVALCANTI; SALES, 2017).

Cada vez mais os pesquisadores são requisitados a depositarem seus dados brutos, ou dados primários, em repositórios de dados de acesso público:

O armazenamento de dados científicos em repositórios e sua reutilização é uma das preocupações do recém-lançado Programa FAPESP de Pesquisa em eScience, expressão que resume o desafio de pesquisa para organizar, classificar e garantir acesso ao gigantesco volume de dados gerados continuamente em todos os campos de pesquisa, a fim de extrair novos conhecimentos e fazer análises abrangentes e originais (MARQUES, 2014, *on-line*).

Ainda sobre o repositório de dados, Sales (2014) afirma que é mais raro no contexto brasileiro. A autora relata que as iniciativas de repositórios de dados, no país, ocorrem no âmbito de consórcios internacionais – por exemplo, o caso do Projeto Genoma Humano. Os repositórios de dados se diferem dos outros repositórios porque o conteúdo, ou melhor, os dados, possuem características próprias, necessitando de um tratamento apropriado, como afirma (RODRIGUES *et al.*, 2010).

Em 2010, o Relatório D-24 do Projeto Repositório Científico de Acesso Aberto de Portugal (RCAAP) relata que as soluções tecnológicas adotadas para repositórios de dados em alguns casos são as mesmas adotadas para outros tipos de repositórios. No documento, é relatado que as agências de fomento à pesquisa tomam a iniciativa de recolher os dados, impulsionadas pelo movimento de acesso livre.

A definição de repositórios de dados de pesquisa apresentada por Costa (2017, p. 46) “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) relatam 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 maneira eficiente.

Devido às peculiaridades de cada disciplina e à variedade e complexidade típica dos resultados de pesquisa, o panorama geral dos repositórios de dados é extremamente heterogêneo em termos de conteúdo, concepção, tratamento de dados e de gestão. Logo, na prática, a escolha da tecnologia a ser aplicada no desenvolvimento do repositório de dados digitais está relacionada à forma como se configurou o processo de curadoria dos dados, a que tipo de dado será armazenado, ao processo de pesquisa (*workflow*) que resultará nos dados e às possibilidades de reuso (SALES, 2014).

Com relação a implantação de repositórios de dados, Sales (2014) afirma que são muitos os benefícios para as comunidades científicas, uma vez que são plataformas padronizados e interoperáveis. De acordo com a autora, tais vantagens são:

Amplia a visibilidade dos resultados de pesquisa, posto que, via de regra, só é formalmente disseminada a fração que está registrada nas publicações acadêmicas. Proporciona mecanismos de preservação de longo prazo, em termos de preservação digital, de arquivamento seguro e de curadoria digital. Permite que o material depositado esteja disponível on-line continuamente para ser consultado e citado mais frequentemente. É um instrumento chave para os processos de reformatação e recriação de dados proporcionados pela curadoria digital. Abre a possibilidade de criação de novos serviços de informação para pesquisadores e gestores a partir da análise dos dados arquivados e a integração de dados e publicações acadêmicas. Permite a criação de redes de repositórios interoperáveis. Aumenta o grau de reuso dos dados minimizando a duplicação de esforços e otimizando os investimentos na geração de dados (SALES, 2014, p. 64).

A quantidade e importância das organizações internacionais envolvidas na padronização, organização e disseminação dos dados de pesquisa reflete a importância dos repositórios de dados de pesquisa (Quadro 13).

Quadro 13 - Organizações com papel chave nos processos de incentivo aos repositórios de dados

INICIATIVA	DESCRIÇÃO
DataCite	Organização sem fins lucrativos formada no final do ano de 2009 em Londres com a responsabilidade de: facilitar o acesso aos dados de pesquisa disponíveis na Internet; aumentar a aceitação de dados de pesquisa como uma contribuição legítima e citável para um registro acadêmico; e dar apoio ao arquivamento de dados de forma a permitir a verificação dos resultados de pesquisa e o reuso dos dados para futuros estudos.
REGISTRY OF RESEARCH DATA RESPOSITORIES (re3data.org)	Fundada pelo <i>German Research Foundation</i> , entre 2012 e 2014, tem como objetivo criar um registro global dos repositórios de dados de pesquisa que cubram diferentes disciplinas acadêmicas, tendo como perspectiva promover a cultura de compartilhamento, aumento do acesso e melhor visibilidade dos dados de pesquisa.
RESEARCH DATA REPOSITORIES - Databib	Define-se como uma ferramenta para apoiar as pessoas a identificarem e a localizarem repositórios on-line de dados de pesquisa. O Databib tem como objetivo responder às seguintes indagações de pesquisadores: quais são repositórios apropriados para um pesquisador submeter seus dados? Como o usuário acha repositórios apropriados e descobre conjunto de dados que se enquadre nas suas necessidades? Como as bibliotecas podem ajudar os usuários a localizar e integrar dados na sua pesquisa ou atividade de ensino?
RESEARCH DATA ALLIANCE-RDA	Tem como objetivo construir pontes técnicas e sociais que permitam o compartilhamento aberto de dados de pesquisa, baseando-se na visão de que pesquisadores e inovadores compartilham livremente dados de forma transversal, sobrepondo-se a tecnologias, disciplinas e países no sentido de equacionar os grandes desafios da sociedade.

Fonte: Sales (2014, p. 65).

De acordo com Sales (2014), os repositórios de dados de pesquisa cumprem missão de construir “pontes de dados” entre disciplinas e domínios fragmentados, que caracterizam

a pesquisa global atual. Ainda segundo a autora, o compartilhamento e a colaboração são chaves para o cumprimento dessa missão, e, além disso, é necessário um conjunto de tecnologias e padrões. O compartilhamento de dados pode ocorrer de forma privada ou informal; pode ser por meio de envio de dados primários, por e-mail, ou outro serviço pessoal. A partir de repositórios, é possível que uma enorme variedade de pesquisadores com os mais diversos níveis de especialização e que estejam espalhados pelo mundo possam atuar no reuso dos dados em suas pesquisas (MEDEIROS; CAREGNATO, 2012).

A criação de repositórios que permitam o reuso de dados em pesquisas futuras possibilita a aceleração de pesquisas a partir da utilização de dados em comum, fazendo com que grandes esforços já despendidos não precisem ser repetidos e permitindo que todos os envolvidos contribuam (MEDEIROS; CAREGNATO, 2012). Sayão e Sales (2012, p. 183) afirmam que várias iniciativas importantes, lideradas pelas próprias comunidades científicas, já cumprem papel vital na garantia do acesso livre aos dados de pesquisa e no que se convencionou chamar de curadoria digital.

Sayão e Sales (2016a, p. 122) discorrem sobre os desafios para a CI, Biblioteconomia e Ciência da Computação em relação aos repositórios digitais de dados de pesquisa, principalmente em termos de prática biblioteconômica e de aplicação de tecnologias. Os autores afirmam que uma das dificuldades é fazer com a diversidade de repositórios, por meio de modelos de dados diferentes, consigam estabelecer:

níveis satisfatórios de interoperabilidade; desenvolver estratégias para a gestão por longo prazo de objetos digitais heterogêneos e complexos – que vão de simples séries numéricas a ambiente de realidade virtual –, o que é muito diferente do que gerenciar publicações digitais, já convencionais, como teses e artigos (SAYÃO; SALES, 2016a, p. 122).

De acordo com Costa (2017, p. 46), é possível analisar a abrangência geográfica e temática dos os repositórios de dados de pesquisa. A autora discorre em sua tese que o país com maior quantidade de repositórios é o Estados Unidos da América, com 40% dos repositórios de dados, seguido pela Alemanha e o Reino Unido, que representam 13% e 11%, respectivamente. No portal do Re3Data sobre repositórios de dados, é possível verificar, atualmente, que essas proporções se mantêm; contudo, vislumbram-se repositórios na América Latina e na África, que representam 1% e 0,8%, respectivamente.

Com relação à abrangência temática, Costa (2017, p. 47) afirma que metade dos repositórios é relacionado às ciências da vida, enquanto a outra metade divide-se em ciências naturais, engenharia, ciências sociais e humanidades. Acredita-se que a predominância de

repositórios ligados à área das ciências da vida ocorre por causa do histórico de compartilhamento de dados. Como citado anteriormente neste estudo, o Projeto Genoma remonta desde a década de 1990, além dos outros estudos que afirmam que tal área compartilha mais dados do que outras.

1.1.1.3 O DataCite

Em 2004, o *German National Library of Science and Technology* (TIB) atribuiu o primeiro *Digital Object Identifier* (DOI) a um conjunto de dados com o objetivo de torná-lo um resultado de pesquisa científica citável. Cinco anos depois, iniciou-se o financiamento do DataCite, um consórcio global que tem o objetivo de atribuir identificadores e DOI para conjuntos de dados e outros resultados de pesquisa. O consórcio foi fundado com o lema “ajudando você a encontrar, acessar e reusar dados” (SAYÃO; SALES, 2012).

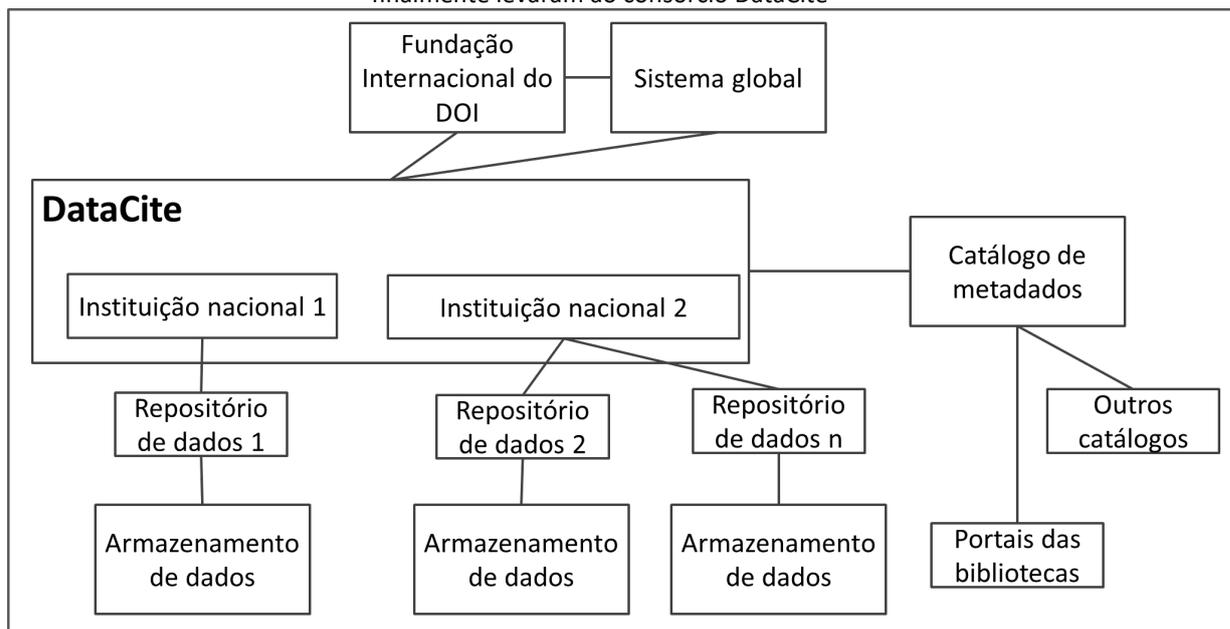
Atualmente, DataCite tornou-se um consórcio global, com 170 membros, distribuídos em mais de 20 países, entre repositório de dados, bibliotecas, agências governamentais, universidades etc., o que atribuiu mais de quatro milhões de DOI a conjuntos de dados científicos e outros resultados de pesquisa (BRASE; LAUTENSCHLAGER; SENS, 2015). De acordo com Sayão e Sales (2012, p. 183), os objetivos essenciais do consórcio são:

[...]estabelecer bases para o acesso mais fácil a dados de pesquisa na internet; aumentar o grau de aceitação dos dados de pesquisa como contribuições legítimas passíveis de serem citadas nos registros acadêmicos; dar sustentação ao arquivamento de dados de pesquisa de forma que seja possível que os resultados possam ser verificados e readaptados para futuros estudos.

O DataCite tem como ideia central a citação de dados, ou seja, os dados de pesquisa devem ser citados da mesma forma como são citadas outras fontes de informação (SAYÃO; SALES, 2012). Os autores defendem que o “princípio da citação de dados permite o reuso e a verificação dos dados mais facilmente, possibilitando que o impacto dos dados possa ser rastreado e que uma estrutura acadêmica que reconheça e recompense os produtores de dados possa ser criada”. Os mesmos autores explanam que:

Para cumprir seus objetivos o DataCite procura juntar as comunidades que lidam com conjunto de dados de pesquisa para que, de forma colaborativa, equacionem o desafio de tornar os dados de pesquisa visíveis e possíveis de serem acessados. Uma das iniciativas importantes nesse processo é o apoio aos centros de dado no assinalamento de identificadores persistentes e na definição de padrões para a publicação de dados; destaca-se também apoio aos editores científicos no sentido de os capacitarem a estabelecer links entre artigos e os dados subjacentes e eles. Para o usuário pesquisador, o DataCite oferece recursos e serviços que o ajudam a encontrar, identificar e citar conjunto de dados de forma confiável (SAYÃO; SALES, 2012, p. 183).

Figura 9 - Estrutura organizacional internacional de agências de publicação e agentes de publicação que finalmente levaram ao consórcio DataCite



Fonte: (BRASE; LAUTENSCHLAGER; SENS, 2015).

Para Neumann e Brase (2014, *tradução nossa*), o DOI vem sendo utilizado como identificador persistente para garantir a citação dos dados de pesquisa. Nesses casos, o DOI é administrado pela Fundação Internacional do DOI (*International DOI Foundation - IDF*). Tal fundação é responsável por “fornecer uma infraestrutura técnica e social para o registro e uso de identificadores interoperáveis persistentes para uso em redes digitais” (INTERNATIONAL DOI FOUNDATION, 2019, *tradução nossa*).

Brase, Lautenschlager e San (2015, *tradução nossa*) afirmam que a seleção do DOI como identificador persistente se deu a partir da analogia entre a publicação de dados de pesquisa e a publicação de literatura científica. Ainda segundo as autoras, outra razão foi a familiaridade dos cientistas com o uso do DOI, pois assim ele tornou-se amplamente utilizado nos artigos de periódicos científicos.

De forma geral, qualquer tipo conteúdo científico em formato digital deveria ser “citável” e, portanto, o uso de um identificador persistente é importante para que se possa encontrar o conteúdo e citá-lo (NEUMANN; BRASE, 2014). O DOI identifica uma entidade em um ambiente digital, ou, em outras palavras, identifica o objeto e não o local (NEUMANN; BRASE, 2014).

O uso do DOI traz benefícios como, por exemplo, economia de tempo e custos – ou seja, por usar um padrão comum, é possível ter maior colaboração e automação; o conteúdo

está sempre rastreável; há a possibilidade de acrescentar informações que permitem a integração com novas ferramentas e com a atualização de informações.

O projeto do DataCite definiu um conjunto de padrões de metadados para o registro do DOI. Este padrão é baseado na ISO 690-2; é um padrão de metadados para descrição bibliográfica. Inclusive os campos destinados à citação têm como padrão adotado o Dublin Core (BRASE; LAUTENSCHLAGER; SENS, 2015).

Para Sales (2014), apenas os padrões de metadados descritivos não são mais suficientes, pois são muitas as questões que devem ser consideradas nos processos de gestão de acervos digitais. A autora relata que diferentes informações precisam estar representadas no objeto para assegurar preservação e confiabilidade das informações nele contidas e, por esse motivo, uma variedade de metadados vem sendo criada em contextos específicos. Destarte, Sales (2014, p. 101) explica que:

Os metadados são utilizados não apenas para descrever o conteúdo do documento, mas também para atribuir identificação persistente, recompor a estrutura do documento, explicitar as relações com outros objetos, com versões, manifestações/expressões, evidenciar declarações de direitos, por exemplo: acesso e *copyright*; informar as dependências técnicas, instruir a preservação, registrar proveniência, isto é, a história do objeto, garantir a integridade e autenticidade do objeto, ou seja, se foram alterados de forma não documentada.

Nesse sentido, o DataCite elaborou um conjunto de metadados. Tal conjunto foi construído alinhando-se aos objetivos do DataCite. Entre eles, cabe destacar que o padrão de metadados aspira recomendar um formato de citação padrão para conjuntos de dados; fornecer a base para interoperabilidade com outros dados; promover a descoberta de conjunto de dados com elementos opcionais, permitindo a descrição flexível do recurso; e estabelecer as bases para serviços futuros (STARR; GASTL, 2011).

Na literatura acerca do tema DataCite, destacam-se alguns estudos sobre a importância do DataCite no compartilhamento de dados. Em 2009, Jan Brase apresenta na *Fourth International Conference on Cooperation and Promotion of Information Resources in Science and Technology*, a pesquisa intitulada *DataCite: a global registration agency for research data*, no qual elucida a criação do consórcio e sua missão de registrar conjuntos de dados de pesquisa e atribuir identificadores persistentes a eles, para que os conjuntos de dados de pesquisa possam ser tratados como objetos científicos únicos, citáveis e independentes.

No ano seguinte, Wilkinson (2010) apresenta o DataCite a comunidade do Reino Unido, uma vez que a Biblioteca Britânica (*British Library*) é a representante do Reino Unido

no DataCite. Em 2011, Starr e Gastl discorrem sobre o esquema de metadados proposto pelo DataCite, destacando o mecanismo para descrever as relações entre o conjunto de dados registrado e outros objetos e, que por ser apoiado organizacionalmente.

Três anos depois, Neumann e Brase (2014) elucidam sobre o DOI e como ele pode ser usado para se referir a dados de pesquisa. Além dos serviços e benefícios para a citação de dados de pesquisa providos pelo DataCite. No ano subsequente, Brase, Lautenschlager e Sens (2015) dissertam sobre os dez anos do uso de DOI para dados científicos e os cinco anos da fundação do DataCite, apresentando o desenvolvimento do consórcio e o esforço cooperativo de sucesso liderado por cientistas, bibliotecários e pesquisadores.

No ano de 2017, há dois trabalhos que demonstram as ações e as lições aprendidas pelo DataCite. O primeiro é o estudo de Rueda, Fenner e Cruse (2017), discorrendo sobre o desenvolvimento de uma série de ferramentas e serviços para promover a adoção da publicação e citação de dados entre a comunidade de pesquisa. Bem como, o trabalho realizado, ao longo dos anos, em colaboração com parceiros interdisciplinares nessas questões, também, as contribuições sobre o desenvolvimento de fluxos de trabalho de publicação de dados. O segundo estudo de Robinson-Garcia *et al.* (2017), explora as características do DataCite para determinar suas possibilidades e potencial como uma nova fonte de dados bibliométricos para analisar a produção acadêmica de dados abertos. Neste artigo, os autores examinam criticamente os registros baixados da API OAI do DataCite e elaboram uma série de recomendações sobre o uso desta fonte para análises bibliométricas de dados abertos, destacando as questões relacionadas à incompletude dos metadados, falta de padronização e definições ambíguas de vários campos.

Por fim, em 2019 o trabalho de Specka *et al.* (2019) utiliza o esquema de metadados do DataCite e outro padrão de metadados para desenvolver um padrão de metadados para dados de pesquisa geoespacial em solo-agrícola. A utilização do padrão de metadados do DataCite, como uma fonte para criação de novo padrão de metadados, específicos para disciplina, reforça o trabalho desenvolvido pelo consórcio para a descrição e interoperabilidade dos dados de pesquisa.

2.3 Sobre a fundamentação teórica

A breve discussão apresentada acerca dos temas deste estudo divide-se em dois elementos. O primeiro relaciona-se com o desenvolvimento da comunicação científica, perpassando os modelos de comunicação, os meios de comunicação com ênfase no periódico

científico, destaque para os *datajournals*, até as discussões sobre Ciência Aberta, com enfoque nos dados abertos. O segundo elemento diz respeito aos fatores envolvidos no processo de comunicação dos dados de pesquisa, definindo o conceito de dado de pesquisa, metadados, princípios FAIR e apresentando o contexto do compartilhamento de dados e repositórios de dados.

O primeiro tópico da fundamentação teórica trata da comunicação científica. Aborda a essência dos processos da comunicação científica para o próprio “fazer ciência”. Em seguida, aponta os modelos de comunicação sistematizados ao longo do tempo, com enfoque nos três processos basilares, uso, produção e compartilhamento de informação científica, enfatizando-se os dados de pesquisa. Dentre os modelos, destaca-se o modelo UNSIST, que desde a década de 1970 apresentava os canais tabulares com a finalidade de comunicar dados técnicos e científicos. A iniciativa possui grande relevância histórica, pois representa a perspectiva visionária da necessidade de considerar dados de pesquisa no fluxo da comunicação da ciência, como relatado por Costa e Leite (2017, p. 91). Na discussão sobre a comunicação científica foi ainda abordado o tema sobre periódicos científicos, visto que, nos modelos de comunicação científicas citados, observa-se a presença do artigo de periódico científico como um canal de comunicação dos resultados de pesquisa. Contudo, na discussão explanou-se sobre os *data journals*, como alternativa a publicação dos dados, canal exclusivo aos artigos de dados (*data paper*), como discorrido por Chavan e Penev (2011).

Ainda no tópico da comunicação científica, discorre-se sobre o tema da Ciência Aberta, pois os seus argumentos motivadores retomam os princípios da ciência moderna, como defendido por Boulton (2013, p. 240, *tradução nossa*). Destacam-se as categorias apresentadas pelo projeto FOSTER (2019), bem como a sua definição de Ciência Aberta.

O segundo aspecto da fundamentação teórica deste estudo tratou especificamente da comunicação dos dados de pesquisa. Primeiramente, definiram-se elementos essenciais para a discussão, tais como o conceito de dado de pesquisa. A definição proposta por Costa (2017) é considerada a mais aproximada para este estudo, isto é, “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”. Neste estudo, ressalta-se que as classificações dos dados foram construídas em contextos específicos, por isso diferem umas das outras, embora atendam aos objetivos pelos quais foram criadas. Portanto, conforme discutido por Torres-Salinas, Robinson-Garcia e Cabezas-Clavijo (2012), as áreas de conhecimento são as responsáveis por estabelecer uma classificação para dados de

pesquisa, porque devem-se considerar os métodos de pesquisa que os geram, as práticas e as normas da área.

Ainda dentro do tópico da comunicação dos dados de pesquisa, apresenta-se o conceito de metadados e a discussão na Ciência da Informação. De forma que, o uso de metadados de qualidade é fundamental para a visibilidade da publicação e sua recuperação no contexto da comunicação científica. Para Bentancourt e Rocha (2012) a qualidade dos metadados atribuídos em sistemas de recuperação da informação precisa ser garantida por meio de políticas de metadados, normas de procedimentos, avaliações e outros. Nesse sentido, assemelha-se as diretrizes propostas pelos FAIR, também discutido neste tópico. Uma vez que, os princípios FAIR são considerados o instrumento norteador para a descoberta, o acesso, a interoperabilidade, o compartilhamento e a reutilização dos dados de pesquisa (WILKINSON, Mark D. *et al.*, 2016).

Em seguida, abordamos o tema do compartilhamento de dados, levantando as questões sobre políticas de compartilhamento, quais existem e quem as desenvolve; também abordamos a curadoria dos dados de pesquisa, visto que o compartilhamento está inserido nesse tema. Assim, como há diferentes definições e classificação para cada área do conhecimento, o mesmo ocorre para o compartilhamento de dados de pesquisa. As áreas do conhecimento se desenvolveram de maneiras diferentes em relação a esse processo. Outro tema abordado são as iniciativas para comunicação dos dados de pesquisa. Salientam-se os 14 princípios e recomendações para a comunicação dos dados elucidados por Costa (2017, p. 72), pois tais princípios e recomendações são significativos para estudar os fundamentos das práticas de comunicação dos dados de pesquisa. De forma complementar, discutem-se os oito fatores que influenciam o compartilhamento de dados de pesquisa apontados por Kim e Stanton (2016), trazendo a perspectiva do indivíduo e institucional, que se complementa, em certa medida, como fatores para o compartilhamento de dados de pesquisa.

Finalmente, o último tópico do contexto de compartilhamento de dados relaciona-se ao repositório de dados abertos, o qual traz a caracterização e as iniciativas a respeito. Destaca-se o consórcio de DataCite, um consórcio global que tem o objetivo de atribuir identificadores e DOI para conjuntos de dados e outros resultados de pesquisa. Sua relevância para estudos sobre compartilhamento de dados pode ser observada nos trabalhos de Wilkinson (2010), Starr e Gastl (2011), Brase (2014), Neumann e Brase (2014), Brase, Lautenschlager e Sens (2015), Robinson-Garcia *et al.* (2017), Rueda, Fenner e Cruse (2017), Datacite Metadata Working Group (2019) e Specka *et al.* (2019). Com destaque, para o

estudo realizado por Robinson-Garcia *et al.* (2017), que assim como este estudo utiliza o recurso de API do DataCite para coleta de metadados. Outro ponto importante, são as parcerias com organizações importantes na discussão de dados abertos de pesquisa e compartilhamento de dados, como, por exemplo, FORCE 11 (*The Future of Research Communications and e-Scholarship*), FAIR Sharing e Research Data Alliance (RDA).

A literatura acerca dos temas pertinentes ao estudo possibilitou analisar, conforme afirmam Sayão e Sales (2012), que a ciência orientada por dados cria um ponto de inflexão no ciclo tradicional da comunicação científica. Observa-se que o conceito de publicação científica pode ser entendido como um conglomerado de produtos científicos, isto é, tornam visíveis tanto os resultados quanto o processo para atingi-los (PAMPEL *et al.*, 2013). No domínio da comunicação dos dados de pesquisa são inúmeras as reflexões que se podem fazer em face dos impactos do compartilhamento de dados de pesquisa, reúso de dados de pesquisa, da publicação e da citação de coleções de dados. De forma geral, a comunicação de dados de pesquisa promove certa velocidade ao ciclo da comunicação científica, na medida em que oferta aos pesquisadores diversos dados, ou seja, dados tratados, que asseguram a fidedignidade de seu significado e a reconstrução correta.

O surgimento de iniciativas que destacam a importância dos dados de pesquisa, entendidos como processos intermediários, define um novo momento na comunicação científica. Pois, se antes apenas importava os resultados da pesquisa, hoje discute-se que, além do produto final dela, os processos intermediários também precisam de visibilidade. A adoção de repositórios de dados como uma modalidade de publicação de dados, para Curty e Aventurier (2017) “oferecem menos atrativos do ponto de vista da lógica do reconhecimento do trabalho científico, considerando que os pesquisadores tendem a preferir os meios e formatos de disseminação científicos legitimados”. Em contrapartida, tais plataformas tem ofertado mais recursos para citabilidade e encontrabilidade dos dados. Diante disso, verificar os dados depositados em repositórios de dados e citados em periódicos científicos, permitem compreender um dos muitos contextos do compartilhamento de dados de pesquisa.

3 METODOLOGIA

Este capítulo apresenta a caracterização da pesquisa do ponto de vista metodológico e os seus procedimentos operacionais. O conjunto desses três elementos compõe a metodologia desta pesquisa e estão baseados nas definições apresentadas por Creswell (2010) e Kauark, Manhães e Medeiros (2010). Conforme definiram Strauss e Corbin (2008), a metodologia é um modo de pensar e estudar a realidade social e, desta maneira, os métodos devem ser sistematicamente organizados, para que seja possível executar e compreender a pesquisa.

3.1 Caracterização da pesquisa

O presente trabalho se caracteriza como uma pesquisa descritiva (KAUARK; MANHÃES; MEDEIROS, 2010, p. 28), pois visa analisar as relações dos conjuntos de dados depositados em repositórios de dados, os quais são membros do consórcio DataCite, com seus artigos de periódicos correspondentes.

Em relação à abordagem do problema, considera-se como uma pesquisa quanti-qualitativa, isto porque o estudo visa coletar e analisar tanto dados quantitativos como qualitativos (CRESWELL, 2010). Empregam-se estratégias de investigação que envolvem coleta das duas formas de dados ao mesmo tempo, durante o estudo, e depois integram-se as informações na interpretação dos resultados gerais (CRESWELL, 2010, p. 33).

A pesquisa possui natureza básica (KAUARK; MANHÃES; MEDEIROS, 2010, p. 26), uma vez que não tem o objetivo gerar conhecimentos para aplicação. Possui, também, horizonte temporal transversal, pois fará análise de uma quantidade de conjuntos de dados em um período específico, porque a coleta na base de dados – que cresce diariamente – deverá ser realizada em dia e horário estabelecidos.

O método adotado foi o indutivo, pois, a partir da identificação dos conjuntos de dados, dos metadados e elementos utilizados para descrição e dos artigos de periódicos correspondentes, foi possível inferir conclusões generalizantes.

O problema de pesquisa exige a definição de um percurso metodológico constituído de critérios que devem ser estabelecidos, como método, natureza, procedimentos técnicos, objetivos e abordagem de coleta de dados.

Os elementos fundamentais para a realização do estudo são os conjuntos de dados que possuem artigos de periódicos relacionados. Obteve-se uma amostra intencional, não

atendendo aos critérios de aleatoriedade. A seleção não probabilística intencional, para Walliman (2005, p. 232), é aquela que se baseia no julgamento do pesquisador ou por acidente, geralmente não sendo utilizada para fazer generalizações sobre toda a população. Dessa maneira, por meio de uso de folhas de estilos (*scripts*) e de solicitações remotas específica via *Application Programming Interface* (API)¹¹ – compatível com o DataCite –, foi possível coletar os conjuntos de dados depositados em repositórios com identificadores persistentes. O uso do recurso API traz vantagem, pois não é necessário o download de bases de dados, ou seja, é possível acessar a base atualizada em tempo real. Importa ressaltar que, para esse estudo, entende-se registro como qualquer arquivo depositado nos repositórios de dados membros do consórcio DataCite.

3.2 Procedimentos metodológicos

Nesta subseção, descrevem-se os procedimentos metodológicos adotados no estudo (Quadro 14). Dividem-se em cinco subseções as etapas percorridas, a saber:

- I. Elaboração do *script*: Descrevem-se quais ferramentas, linguagens e bases consultadas;
- II. Seleção da amostra: Apresentam-se os procedimentos adotados para seleção da amostra do estudo com base da coleta via *script*;
- III. Análise do compartilhamento de dados pelas diferentes áreas do conhecimento: Relatam-se os métodos para análise do objetivo OE1 deste estudo;
- IV. Análise do nível de descrição dos registros de conjunto de dados: Relatam-se os métodos para análise do objetivo OE2 deste estudo;
- V. Roteiro de avaliação das políticas de compartilhamento de dados: Relatam-se os métodos para análise do objetivo OE3 deste estudo.

¹¹ Em português, Interface de Programação de Aplicações. Trata-se de um conjunto de padrões e rotinas estabelecidas que permite a terceiros usarem seus serviços por meio da web.

Quadro 14 - Procedimentos metodológicos adotados no estudo conforme objetivos específicos

Objetivo Específico	Fonte	Coleta de dados	Análise de dados
Compreender como as diferentes áreas do conhecimento compartilham dados	Amostra	<i>Script</i> (via API)	Análise estatística
Analisar o nível de descrição registros de dados coletados pelas diferentes áreas do conhecimento	Amostra Site dos registros de conjunto de dados	<i>Script</i> (via API) Lista de verificação	Análise estatística Análise textual
Avaliar as políticas de depósito de dados de pesquisa de periódicos com o intuito de verificar a relação entre as políticas e as áreas do conhecimento	Amostra Site dos periódicos	<i>Script</i> (via API) Lista de verificação	Análise estatística Análise textual

Fonte: Elaboração da autora (2019).

3.2.1 Elaboração do script

Para a coleta de dados foram elaboradas folhas de estilo (*scripts*) escritas em linguagem *Python* por meio da ferramenta *Jupyter Notebook*. Essa ferramenta é de narrativa computacional. De acordo com Perez Granger (2014), “A missão do *Project Jupyter* é criar ferramentas de código aberto para computação científica interativa e a ciência de dados no que tange à pesquisa, à educação e à indústria, com ênfase em usabilidade, colaboração e reprodutibilidade”. Portanto, pode-se dizer que a ferramenta está alinhada à filosofia *open*. Além do mais, a ferramenta é bastante elogiada pelos cientistas de dados, pois oferta funcionalidades simples e efetivas (PERKEL, 2018).

Para a coleta de dados foi selecionada a base do DataCite. O DataCite é uma organização global sem fins lucrativos que fornece identificadores persistentes (DOIs especificamente) para dados de pesquisa e outros resultados. As comunidades científicas juntam-se ao DataCite como membros para poder atribuir DOIs a todos os resultados de pesquisa. Assim, todos os resultados de pesquisa tornam-se detectáveis e os metadados associados são disponibilizados à comunidade. Atualmente, o DataCite possui 170 membros, distribuídos em mais de 20 países, entre repositório de dados, bibliotecas, agências governamentais, universidades etc. O site do DataCite fornece documentação e informações sobre o esquema de metadados e como realizar a coleta via API, fatores que corroboram para a elaboração do *script*.

Além do DataCite, outra fonte consultada foi a organização CrossRef. Fundada por editores científicos no ano de 2000, a Crossref surgiu da necessidade de fornecer um serviço colaborativo central de referência – usando DOIs – que seria acessível e gerenciado pela comunidade global de comunicação científica (WOOD, 2018). Atualmente, a CrossRef é uma organização sem fins lucrativos autossustentável que visa tornar todos os resultados de

pesquisa não apenas fáceis de encontrar, citar e vincular, mas também fáceis de avaliar e reutilizar (WOOD, 2018).

Outra fonte consultada foi o *Directory of Open Access Journals* (DOAJ) ou Diretório de Revistas de Acesso Aberto, lançado em 2003 na Universidade de Lund, na Suécia, com 300 revistas de acesso aberto. Atualmente, contém cerca de 12000 revistas de acesso aberto, cobrindo todas as áreas conhecimento, desde ciência, tecnologia, medicina, ciências sociais e humanidades (DOAJ, 2019). Sua missão é aumentar a visibilidade, acessibilidade, reputação, uso e impacto de periódicos de pesquisa acadêmica de qualidade, revisados por pares e de acesso aberto em todo o mundo, independentemente da disciplina, geografia ou idioma (DOAJ, 2019). É uma organização sem fins lucrativos gerenciada pelo *Infrastructure Services for Open Access C.I.C.* com sede no Reino Unido, e depende inteiramente de doações voluntárias de seus membros e de verbas de patrocínio recebidas.

Etapa 01 – Recorte temporal da base de dados

O recorte temporal foi de 12 anos, isto é, foram extraídos os registros depositados entre os anos de 2008 e 2019. Contudo, o objetivo do estudo foi investigar os registros depositados em repositórios apenas entre os anos 2009 e 2019. Isto porque oficialmente, conforme apontado na literatura, o DataCite foi criado em 2009. Porém, por problemas com a API, ao delimitar o recorte no ano de 2009, não foi possível coletar nenhum registro desse ano. Portanto, foram extraídos do *script* registros com o campo *dates* (data) preenchidos de 2008 a 2019.

Um *script* auxiliar foi elaborado para tratar o campo *dates*, porque não há um padrão de entrada de datas. Assim sendo, recuperaram-se valores do tipo: aaaa, dd-mm-aaaa, aaaa-mm-dd etc. Este *script* auxiliar foi elaborado para contar quantos caracteres existiam em cada campo *dates*. O objetivo final era que, após sua execução, o campo *dates* ficasse padronizado no formato aaaa.

Tabela 1 - Quantidade de caracteres no campo *dates* por registro

Quantidade de caracteres	Quantidade de registros
5	22247
8	16
9	1
11	10576
13	24

Quantidade de caracteres	Quantidade de registros
19	1
22	2
25	8
26	24
Vazio	346
Total	33245

Fonte: Elaboração própria (2019).

Etapa 02 – Definição dos metadados para coleta no DataCite

O segundo momento foi delimitar quais metadados seriam consultados. Utilizou-se o esquema de metadados do DataCite para a seleção dos campos, de acordo com a documentação disponibilizada no website do consórcio. Tal esquema é baseado no padrão DublinCore, uma vez que a organização apoia e colabora com *Dublin Core Metadata Initiative* (DCMI). Existem três níveis diferentes de obrigação para as propriedades de metadados:

- a) Propriedades obrigatórias (M) devem ser fornecidas obrigatoriamente;
- b) Propriedades recomendadas (R) são opcionais, mas altamente recomendadas para interoperabilidade; e
- c) Propriedades opcionais (O) são opcionais e fornecem uma descrição mais rica.

Optou-se por recuperar todos metadados obrigatórios (M). Por isso, selecionaram-se os seguintes metadados: *Identifier*, *Creator*, *Title*, *Publisher*, *PublicationYear*, *ResourceType*. Em consonância com o objetivo deste estudo, fez-se necessária a recuperação de metadados responsáveis por estabelecer a relação entre os registros depositados nos repositórios de dados e os seus respectivos documentos relacionados – no caso estudado, artigos de periódicos.

Por esse motivo, optou-se por selecionar o metadados *relationType*. É um metadado de preenchimento opcional, mas quando preenchido indica se o dado é parte de ou foi citado por outro documento. Esses metadados representam um campo não repetitivo e seu preenchimento é por meio de uma lista com 34 valores autorizados. Desses valores, selecionaram-se quatro; *IsCitedBy*, *Cites*, *IsSupplementTo*, *IsSupplementBy*, *IsPartOf*,

HasPart. Os valores foram escolhidos, pois fazem parte do “superconjunto”, isto é, o DataCite definiu que a combinação de alguns campos, obrigatórios e recomendados, aprimora a recuperação, citação e links dos registros.

Quadro 15 - Metadados que indicam relacionamento e são parte do "superconjunto"

IsCitedBy	indica que B inclui A em uma citação
Cites	indica que A inclui B em uma citação
IsSupplementTo	indica que A é um complemento para B
IsSupplementBy	indica que B é um complemento para A
IsPartOf	indica que A é uma porção de B; pode ser usado para elementos de uma série
HasPart	indica A inclui a parte B

Fonte: Traduzido de Datacite Metadata Working Group (2019).

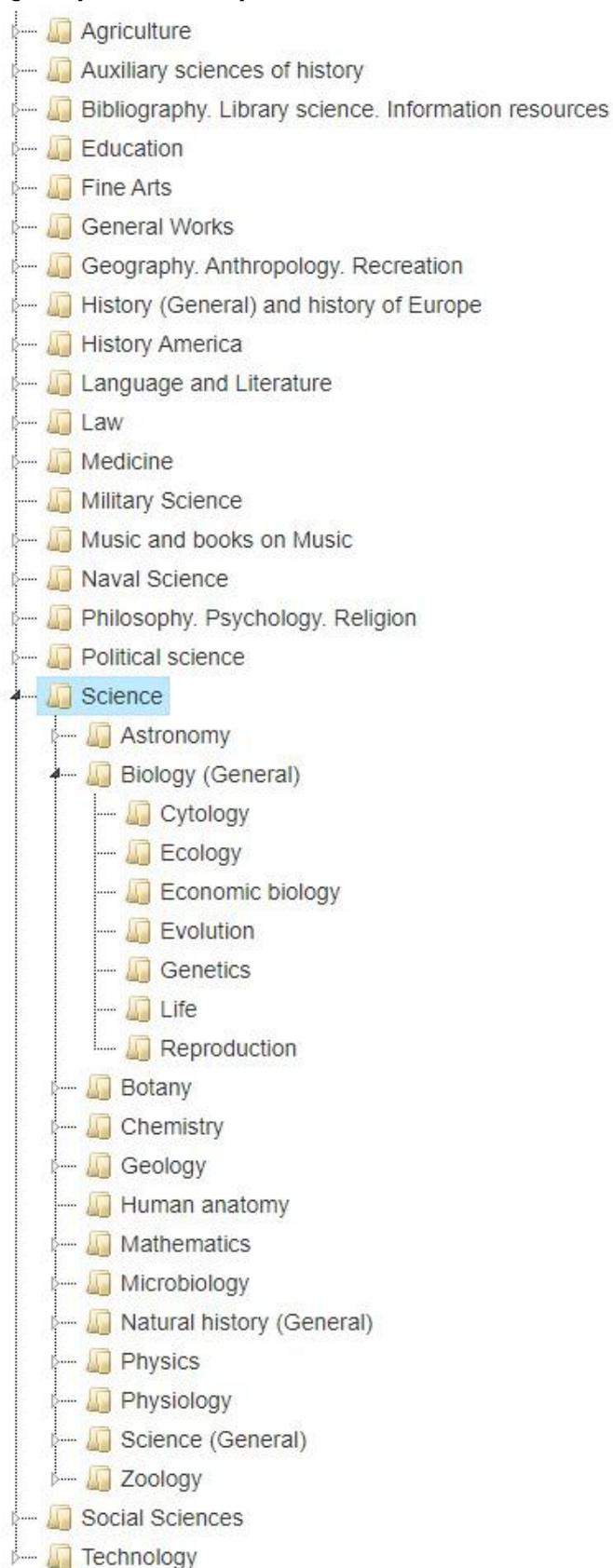
Por fim, o último campo é *relatedIdentifierType*. Ele também faz parte do “superconjunto” e seu preenchimento ocorre de acordo com a lista de valores autorizados, esses são: ARK, arXiv, bibcode, DOI, EAN13, EISSN, Handle, IGSN, ISBN, ISSN, ISTC, LISSN, LSID, PMID, PURL, UPC, URL, URN e w3id. Optou-se por utilizar o identificador persistente DOI, como descrito na literatura. Ele vem sendo utilizado como identificador persistente por garantir a citação dos dados de pesquisa.

Etapa 03 – Consulta à agência CrossRef e ao DOAJ

Como nem todo DOI coletado do metadado *relatedIdentifierType* corresponde a um artigo de periódico, realizou-se consulta na agência CrossRef para identificar quais dos documentos eram artigos de periódicos e quais não. A partir da consulta a CrossRef foi possível coletar os metadados referentes ao ISSN, Título do artigo e a Editora responsável. Assim sendo, a presença do campo ISSN na amostra funcionou como uma espécie de “validador” para coletar os registros de conjunto de dados, isto é, quando o campo do ISSN está preenchido, o registro é selecionado para a amostra. Entretanto, cabe ressaltar que o *script* apenas coletou os metadados, e a verificação do campo ISSN foi realizada na seleção da amostra, descrita na próxima subseção.

A partir da coleta do metadado com o ISSN dos periódicos foi possível consultar se essas revistas científicas estavam presentes no DOAJ, pois, baseado na literatura específica, o compartilhamento de dados também está relacionado com a ideologia do pesquisador, isto é, pesquisadores que apoiam a filosofia aberta tendem a publicar em periódicos abertos e a compartilhar seus dados mais dos que os outros pesquisadores (ANDREOLI-VERSBACH; MUELLER-LANGER, 2014; PIWOWAR, 2011). Para a verificação de qual área do conhecimento ea que revista pertence, coletaram-se o código da área do conhecimento e a área de conhecimento definida pelo DOAJ.

Figura 10 - Organização da classificação de área do conhecimento do DOAJ



Fonte: Elaboração própria (2019).

A classificação das áreas do conhecimento do DOAJ se organiza em 20 classes. Tais classes subdividem-se em subclasses mais específicas, por exemplo, a classe *Science* divide-se em 13 subclasses, que também se dividem em subclasses, e assim por diante. Importa ressaltar que a classificação adotada pelo DOAJ é a desenvolvida pela *Library of Congress* (LOC).¹²

3.2.2 Seleção da amostra

Após a elaboração e execução do *script*, os dados foram exportados para uma planilha no formato .csv. Totalizaram-se 33.245 registros, de modo que a quantidade de registros por ano não foi igual porque o script realiza a coleta aleatoriamente (Quadro 17).

Tabela 2 - Lista de registros coletados do DataCite por ano

Ano	Quantidade de registros
Vazio	347
2008	3221
2009	921
2010	1247
2011	637
2012	870
2013	1402
2014	1600
2015	4739
2016	4192
2017	4967
2018	5148
2019	3954
Total	33245

Fonte: Elaboração própria (2019).

Em consonância com o objetivo de analisar o compartilhamento de dados, optou-se por aplicar o recurso do filtro, disponibilizado pelo editor de planilhas na coluna *resourceTypeGeneral* (indica o tipo de material). De acordo com o objeto deste estudo, selecionou-se o tipo *datasets* (conjunto de dados), totalizando 16.279 registros.

¹²Disponível em: <https://www.loc.gov/catdir/cpsolcco/>

Tabela 3 - Lista do tipo de material coletado no DataCite por registros

Metadado resourceTypeGeneral	Quantidade de registros.
Collection	2400
Dataset	16282
Text	11947
Image	1338
InteractiveResource	208
Audiovisual	129
Other	224
Software	69
article	314
Sound	150
Event	131
Service	27
Model	6
PhysicalObject	17
Vazio	3
Total	33245

Fonte: Elaboração própria (2019).

Desses registros do tipo *datasets*, selecionaram-se somente aqueles que continham o campo ISSN preenchido, totalizando 1.312 registros. Por fim, aplicou-se o filtro para identificar quais revistas estavam cadastradas no DOAJ, totalizando 68 registros.

Apesar do *script* coletar a classificação do DOAJ, para esta pesquisa optou-se por utilizar a Classificação Internacional de Educação (ISCED).¹³ Portanto, foi necessário realizar a correspondência entre a classificação adotada pelo DOAJ e a ISCED.

Com relação à ISCED, trata-se de uma nomenclatura internacional da Organização das Nações Unidas para a Educação, a Ciência e a Cultura (Unesco), destinada à área de Ciência e Tecnologia (UNESCO, 2012). A adoção desta classificação ocorre porque há, na literatura sobre a comunicação dos dados de pesquisa, o compartilhamento de dados e dos repositórios de dados, estudos qualitativos detalhados e pesquisas mais amplas avaliando práticas, desafios e incentivos de compartilhamento de dados entre pesquisadores. Alguns dos estudos se concentram em disciplinas específicas de pesquisa, enquanto outros analisam várias disciplinas. Assim sendo, optou-se por utilizar uma nomenclatura internacional no intuito de identificar as áreas do conhecimento, tendo-se em conta que o estudo se concentra em conjuntos de dados e advém de trabalho relacionado à pesquisa/educação dentro de

¹³ Em inglês: *International Standard Classification of Education*

programas de pós-graduação e não ciência pura, como a desenvolvida em empresas, por exemplo. Outro fator que estimula a utilização da ISCED é que ela pertence ao Grupo Internacional das Nações Unidas de Classificações Econômicas e Sociais, e é aplicada mundialmente em estatísticas, com o objetivo de reunir, compilar e analisar dados comparáveis internacionalmente. A ISCED é a classificação de referência para a organização de programas de educação e qualificações relacionadas por níveis e campos de educação (UNESCO, 2012). Foi projetada para servir como uma estrutura internacional de classificação para as atividades educacionais. Portanto, as definições e os conceitos básicos da ISCED são válidos internacionalmente, além de abrangerem toda a gama de sistemas educacionais.

Com relação à correspondência entre a classificação do DOAJ e a ISCED, foi adicionada uma coluna denominada “Classificação” na planilha identificada como Amostra. Nessa coluna, foi incluída a classificação ISCED da Unesco. Como ambas as classificações apresentam os termos em inglês, realizou-se a busca dos termos nessa língua.

Todos os periódicos foram classificados a partir da primeira subdivisão de assunto, isto é, a classificação IESCD apresenta nove classes gerais de assunto, representadas por um dígito numeral (excluindo o número zero), que se subdividem em classes específicas de assunto, representadas por dois dígitos numerais, excluindo o número zero, à esquerda (Figura 10).

Figura 11 - Classe geral de assunto da ISCED

05 Natural Sciences, Mathematics and Statistics

051 Biological and related sciences

0511 Biology

Biology is the study of the structure, function, reproduction, growth, all living organisms.

Programmes and qualifications with the following main content are c

Biology
Botany
Cell biology
Entomology
Genetics
Mycology
Zoology

Fonte: Captura de tela (2019).

Durante esse processo de classificação, observou-se que a revista *Royal Society Open Science* (2054-5703) possui caráter multidisciplinar, isto é, o mesmo fascículo publica artigos de diferentes áreas do conhecimento. Portanto, para identificar a classificação da revista, foi necessário acessar os 11 artigos coletados via *script* e identificar quais eram as áreas do conhecimento, de modo que nove artigos eram da área *Life Sciences* (Ciência da vida), um da *Physics* (Física) e o outro da *Psychology* (Psicologia) (Quadro 19).

Quadro 16 - Correspondência entre a classificação do DOAJ e a Nomenclatura da Unesco para a área de Ciência Tecnologia (ISCED)

	NOME DO PERIÓDICO	ISSN	CLASSIFICAÇÃO DOAJ	NOMENCLATURA DA UNESCO
1.	College and Research Libraries	0010-0870	Bibliography. Library science. Information resources	Journalism and information
2.	Food Science and Technology	0101-2061	Nutrition. Foods and food supply	Agriculture
3.	ZooKeys	1313-2989	Zoology	Biological and related sciences
4.	Biodiversity Data Journal	1314-2836	Biology (General)	Biological and related sciences
5.	Croatica Chemica Acta	1334-417X	Chemistry	Physical sciences
6.	Kidney & Blood Pressure Research	1423-0143	Dermatology	Health

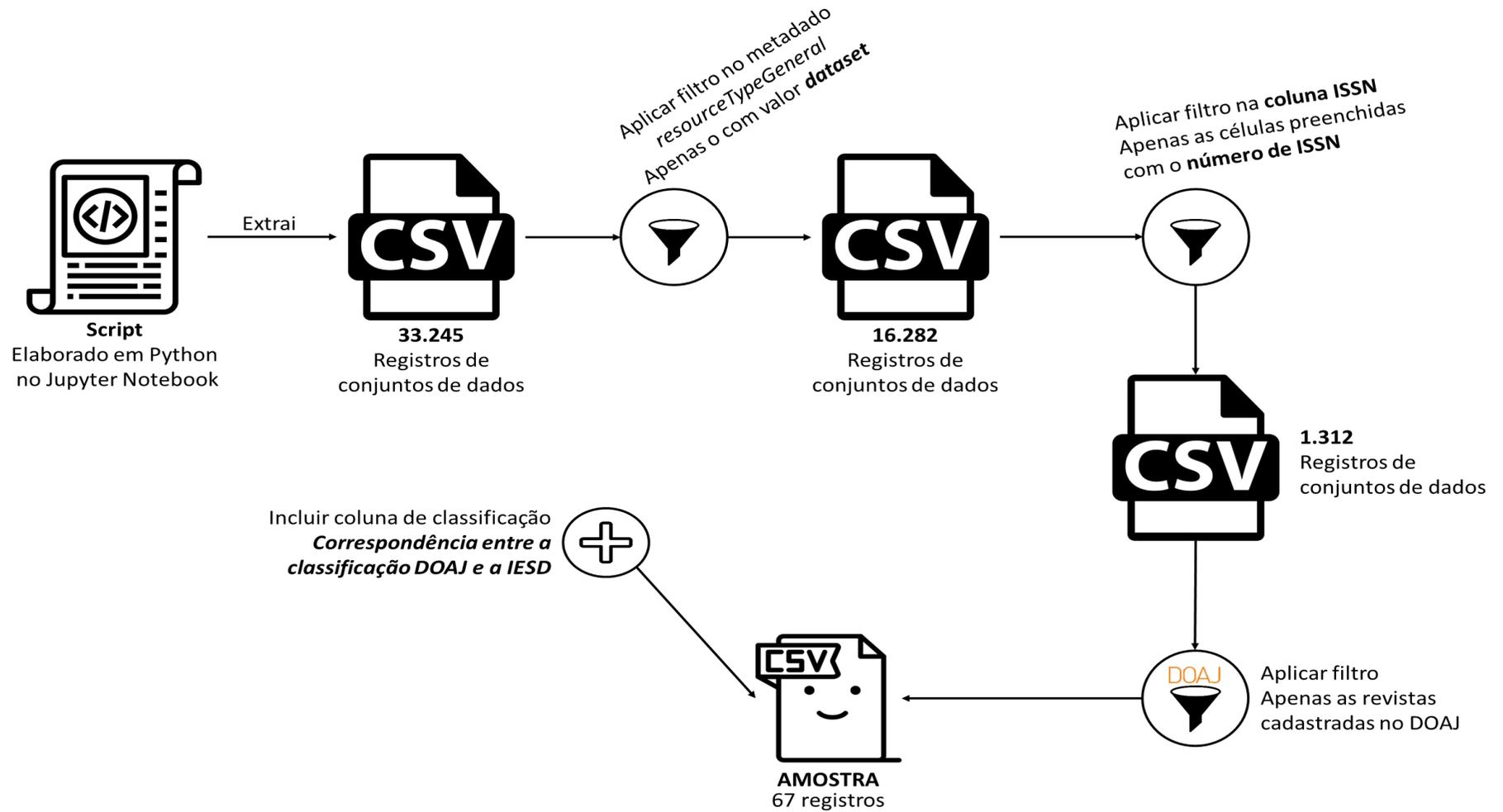
	NOME DO PERIÓDICO	ISSN	CLASSIFICAÇÃO DOAJ	NOMENCLATURA DA UNESCO
7.	Zeitschrift für Kristallographie - New Crystal Structures	1433-7266	Crystallography	Physical sciences
8.	Bioinorganic Chemistry and Applications	1687-479X	Biotechnology	Biological and related sciences
9.	Molecular Systems Biology	1744-4292	Evolution	Biological and related sciences
10.	Clinical Epigenetics	1868-7083	Medicine	Health
11.	Rice	1939-8433	Plant culture	Biological and related sciences
12.	BioTechniques	1940-9818	Biology (General)	Biological and related sciences
13.	Chemical Science	2041-6539	Chemistry	Physical sciences
14.	Ecology and Evolution	2045-7758	Ecology	Biological and related sciences
15.	Royal Society Open Science	2054-5703	Science	Biological and related sciences
16.	Royal Society Open Science	2054-5703	Science	Physical sciences
17.	Royal Society Open Science	2054-5703	Science	Social and behavioural sciences
18.	Earth and Space Science	2333-5084	Astronomy	Physical sciences

Fonte: Elaboração da autora (2019).

Como observado no Quadro 19, a revista *Royal Society Open Science* (2054-5703), por conter artigos em mais de uma área do conhecimento, aparece mais de uma vez na listagem de periódicos selecionados para análise. Para fins de exame, quando se considerar a área do conhecimento, será considerado o total de 18 periódicos e não 16.

Durante o processo de classificação, observou-se que um registro estava duplicado. Apesar de existir dois DOIs para cada registro de conjunto de dados, eles estavam direcionando para o mesmo registro no repositório. Por isso, como representava o mesmo registro de conjunto de dados, optou-se por retirá-lo da amostra. Cabe ressaltar que os registros dos repositórios também indicam área de conhecimento, mas a área do conhecimento utilizada neste estudo foi a coletada do DOAJ, advinda da revista científica e classificada de acordo com a ISCED. Por fim, a amostra de registros de conjunto de dados depositados em repositórios foi de 67. Na Figura 11 apresenta-se, de maneira resumida, os processos para a seleção da amostra deste estudo.

Figura 11 - Procedimentos metodológicos



Fonte: Elaboração própria (2019).

3.2.3 *Análise do compartilhamento de dados pelas diferentes áreas do conhecimento*

Para atingir o objetivo específico, que é compreender como as diferentes áreas do conhecimento compartilham dados, adotaram-se técnicas estatísticas descritivas para a análise. A estatística descritiva é a etapa inicial da análise utilizada para descrever e resumir os dados (DAVILA, 2007, p. 5). Assim sendo, a partir da interpretação dos dados contidos na amostra deste estudo, foi possível extrair as informações acerca do compartilhamento pelas áreas do conhecimento. Para tanto, utilizou-se o software Excel para manipular os dados – por exemplo, o recurso Tabela Dinâmica para cruzar os dados da amostra e a criação de gráficos para visualização dos resultados.

A primeira técnica utilizada foi a frequência absoluta, isto é, o número de vezes em que uma determinada variável assume um valor (AZEVEDO, 2016). No caso do estudo, as variáveis observadas foram: áreas do conhecimento, metadado do tipo de relação e ISSN. Outra técnica adotada foi a distribuição da frequência. Ela foi aplicada anualmente para análise dos quantitativos de registros de dados em cada disciplina.

1.1.2 *Análise do nível de descrição dos registros de conjunto de dados*

Para atingir o objetivo específico de analisar o nível de descrição dos registros de dados coletados pelas diferentes áreas do conhecimento elaborou-se uma lista de verificação para comparar os metadados utilizados nos registros de dados. Como princípio orientador para descrição de conjuntos de dados, optou-se por seguir os apontados pelos os princípios FAIR. Os princípios FAIR, conforme discorrido na literatura, são importantes, pois descrevem como os resultados da pesquisa devem ser organizados para ser acessados, compreendidos, trocados e reutilizados com mais facilidade. Além disso, optou-se por utilizar também as recomendações do DCC, um centro de especialização reconhecido internacionalmente em curadoria digital, com foco na criação de capacidades e habilidades para o gerenciamento de dados de pesquisa.

De acordo com DCC, muitas áreas do conhecimento apoiaram iniciativas para formalizar padrões de metadados, pois eles são necessários para a reutilização de dados. Para as disciplinas que ainda não estabeleceram um padrão de metadados e para os repositórios que trabalham com dados em várias disciplinas, o DCC separou uma seção intitulada Dados Gerais de Pesquisa (*General Research Data*). O objetivo é informar sobre padrões mais amplos de metadados que foram adaptados para atender às necessidades dos dados de

pesquisa. Entre os padrões sugeridos está o DataCite e, como a amostra estudada deriva do DataCite, optou-se por utilizar esse padrão para análise de descrição dos registros.

O *script* elaborado para a coleta dos dados deste estudo já atende a algumas das recomendações dos princípios FAIR, pois foram coletados conjuntos de dados de repositórios membros do DataCite e, portanto, todos conjuntos possuem obrigatoriamente um identificador persistente. Outro metadado coletado é o que indica o relacionamento entre o seu conjunto e o artigo de periódico, atendendo ao critério de incluir referências qualificadas para outros metadados. Além desses dois pontos, a amostra atende aos princípios F4 e A1, os quais orientam, primeiro, que os metadados devem ser indexados e, segundo, que os mesmos devem utilizar protocolos de comunicação. E uma vez que são registros de repositórios membros do DataCite, esses princípios são atendidos, pois constituem parte do serviço do consórcio.

Nesse sentido, a lista de verificação (Quadro 20) foi elaborada não a fim de sistematizar os metadados que foram coletados pelo *script*, mas os utilizados na descrição dos registros depositados nos repositórios. Para a elaboração dessa lista de verificação utilizaram-se as recomendações dos princípios FAIR, que selecionaram os campos que deveriam ser examinados no intuito de analisar a descrição dos registros em conformidade com tais princípios. Para a análise, optou-se por atribuir um número de 1 a 67, com vistas a identificar os registros listados em ordem alfabética de repositório.

Quadro 17 - Lista de verificação de metadados

<i>Número do registro</i>	
Possui campo 'Como citar'?	SIM/NÃO
Possui campo que indica a versão ou data de depósito e/ou alteração?	SIM/NÃO
Apresenta indicação de licença de uso?	SIM/NÃO
Apresenta uso de vocabulário que seguem os FAIR?	SIM/NÃO
Possui campos que descrevem como os dados foram coletados? Ou como e onde foram gerados?	SIM/NÃO

Fonte: Elaboração própria (2019).

A partir da elaboração da lista de verificação optou-se por realizar a captura da tela de cada registro no seu respectivo repositório com a finalidade de se poder realizar a análise, mesmo que o repositório passe por alguma dificuldade técnica ou fique fora do ar. Dessa maneira, no dia 27 de dezembro de 2019, foram realizadas as capturas de tela utilizando o recurso disponibilizado pelo próprio computador. Em seguida, para análise, optou-se por atribuir notas de acordo com a resposta: a cada SIM recebido é atribuído 1 ponto, para cada NÃO é atribuído 0. Logo, caso o registro possua todos os campos, receberá a nota 5.

3.2.4 Roteiro de avaliação das políticas de compartilhamento de dados

Além dos artigos de periódicos, o *script* elaborado coletou as suas respectivas revistas, isto é, trouxe os metadados responsáveis pelo ISSN. Assim, por meio de navegação nas páginas web – ou seja, observação direta –, buscou-se identificar a presença de políticas de compartilhamento de dados nas revistas da amostra. No total, foram identificadas 16 revistas de diferentes áreas do conhecimento para análise.

Em relação à política de compartilhamento de dados, procurou-se averiguar se existia uma política e, em casos positivos, se era uma política do tipo que encorajava ou obrigava o compartilhamento dos dados. A análise foi realizada a partir da leitura das políticas editoriais de cada uma das revistas. Para fins de controle, elaboraram-se fichas, como exemplificado no Quadro 21.

Quadro 18 - Ficha de avaliação das revistas

Nome da revista:		
Possui política de compartilhamento de dados?	SIM	NÃO
Se sim, a política encoraja ou obriga o compartilhamento de dados em repositórios?	OBRIGA	ENCORAJA
Há indicação de repositórios de dados para o depósito?	SIM	NÃO
Se não, há menção da importância de citar os dados utilizados ou inclusão como suplemento no processo de submissão?	SIM	NÃO

Fonte: elaboração própria (2019).

Para encontrar as políticas de compartilhamento de dados, os termos em inglês utilizados foram: *data policy*, *policy*, *data availability statement* e *sharing data policy*. Os termos em inglês para identificar se as políticas obrigam ou encorajam foram palavras que indicam obrigação: *must*, *obligate* e *require*; e palavras que encorajam: *should*, *strongly recommend* e *reccomend*.

4 RESULTADO E DISCUSSÕES

Esta seção tem por objetivo apresentar e discutir os resultados da pesquisa proposta por esta dissertação. A investigação tem como foco principal analisar a relação dos conjuntos de dados depositados em repositórios com os periódicos científicos. Cabe ressaltar, que a coleta dos dados foi realizada dia 26 de novembro de 2019, e a base do DataCite está em crescimento contínuo, no dia da extração dos dados, a base apresentava aproximadamente 16 milhões de registros.

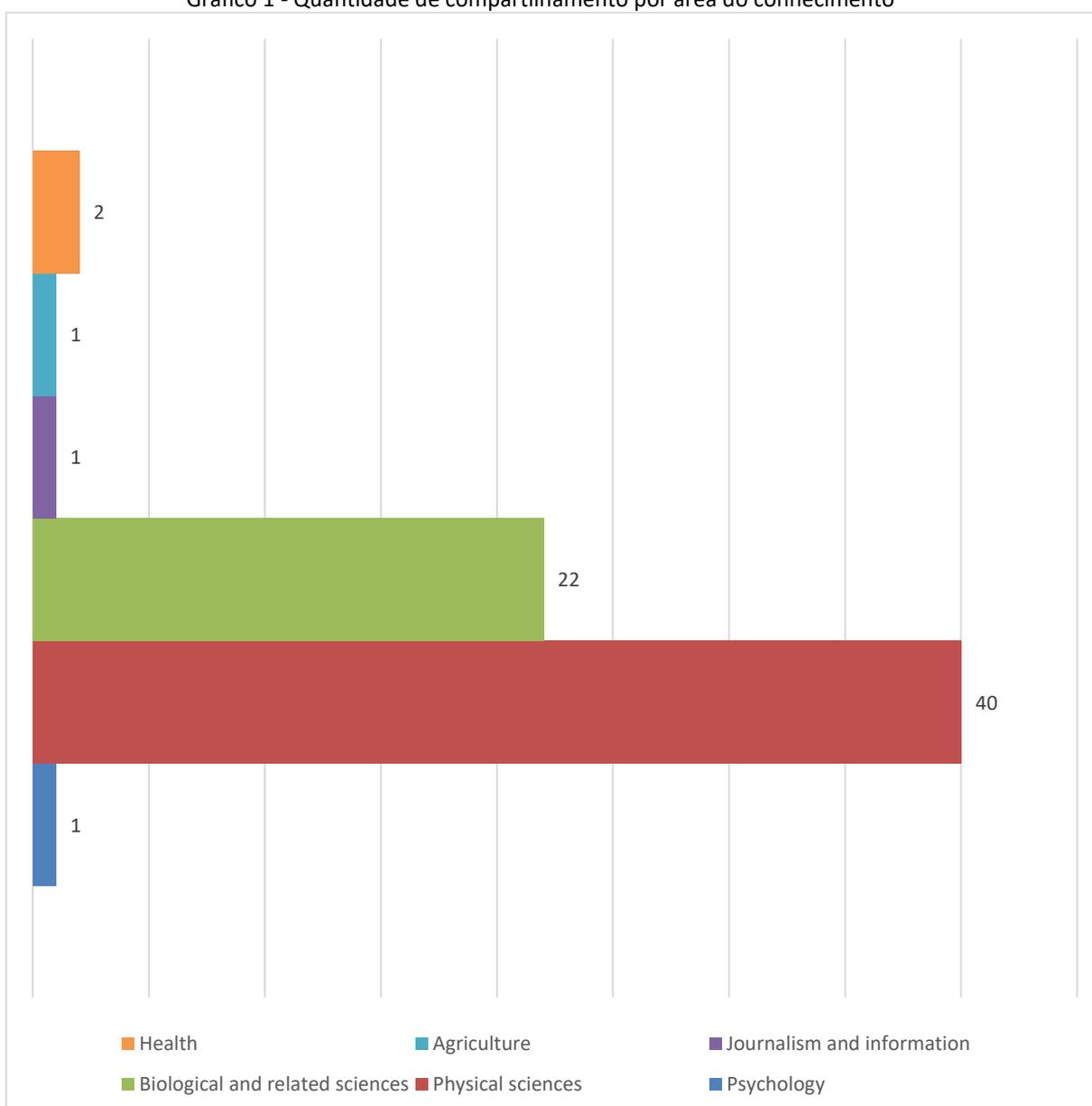
4.1 Compartilhamento de dados nas áreas do conhecimento

A amostra analisada possui 67 registros de conjuntos de dados vinculados a artigos de periódicos. Esses artigos pertencem a 16 títulos de periódicos. Dessa amostra, constatou-se que 64 registros (96%) utilizaram o metadado *IsSupplementTo* (É suplemento de/para) para relacionar o registro de conjunto de dados ao artigo de periódico, isto é, os dados sugerem que o seu produtor é o mesmo autor do artigo, indicando ainda que os pesquisadores estão compartilhando seus dados. Enquanto isso, apenas três registros (4%) utilizaram o metadado *IsCitedBy* (Citado por), após navegação nas páginas dos registros de conjunto de dados, identificou-se que, na verdade, são dois registros de conjunto de dados e não três. Pois, dois registros, identificados na amostra como registro nº 40 e 41, correspondem ao mesmo conjunto de dados, tal verificação é constatada por meio do identificador DOI, que é igual nos dois registros. Todavia, esse registro de conjunto de dados é citado em dois artigos científicos diferentes.

Em relação aos registros nº40 e 41, define-se neste estudo como **Registro A** para de facilitar a análise e a discussão. Após a navegação na página do Registro A, constata-se que o conjunto de dados foi citado por oito artigos científico diferentes, sugerindo o reúso do conjunto de dados. Ainda sobre o reúso de dados, no terceiro registro de conjunto de dados, identificado na amostra como registro nº66, identifica-se que o conjunto de dados foi citado em 24 estudos diferentes. Dentre esses 24, verifica-se que um dos estudos é também parte do estudo que gerou o conjunto de dados, isto é, uma pesquisa gerou como resultado de pesquisa um artigo científico e um conjunto de dados. Portanto, observa-se o reúso de dados de pesquisa, uma vez que estão depositado em repositórios de dados e possuem identificadores persistentes.

Na amostra, nota-se a presença de seis áreas do conhecimento: *Journalism and information* (Jornalismo e informação), *Agriculture* (Agricultura), *Biological and related sciences* (Biologia e áreas correlatas), *Physical sciences* (Ciências físicas), *Health* (Saúde) e *Social and behavioural sciences* (Ciências sociais e comportamentais). Entre as revistas, a área do conhecimento denominada *Physical sciences* (Ciências físicas) teve o maior número de compartilhamentos e/ou reuso de dados, seguida por *Biological and related sciences* (Biologia e áreas correlatas) (Gráfico 1). Como apontado por Costa (2017) em sua tese, entre vários fatores, a área do conhecimento influencia a abertura ou não dos dados de pesquisa.

Gráfico 1 - Quantidade de compartilhamento por área do conhecimento

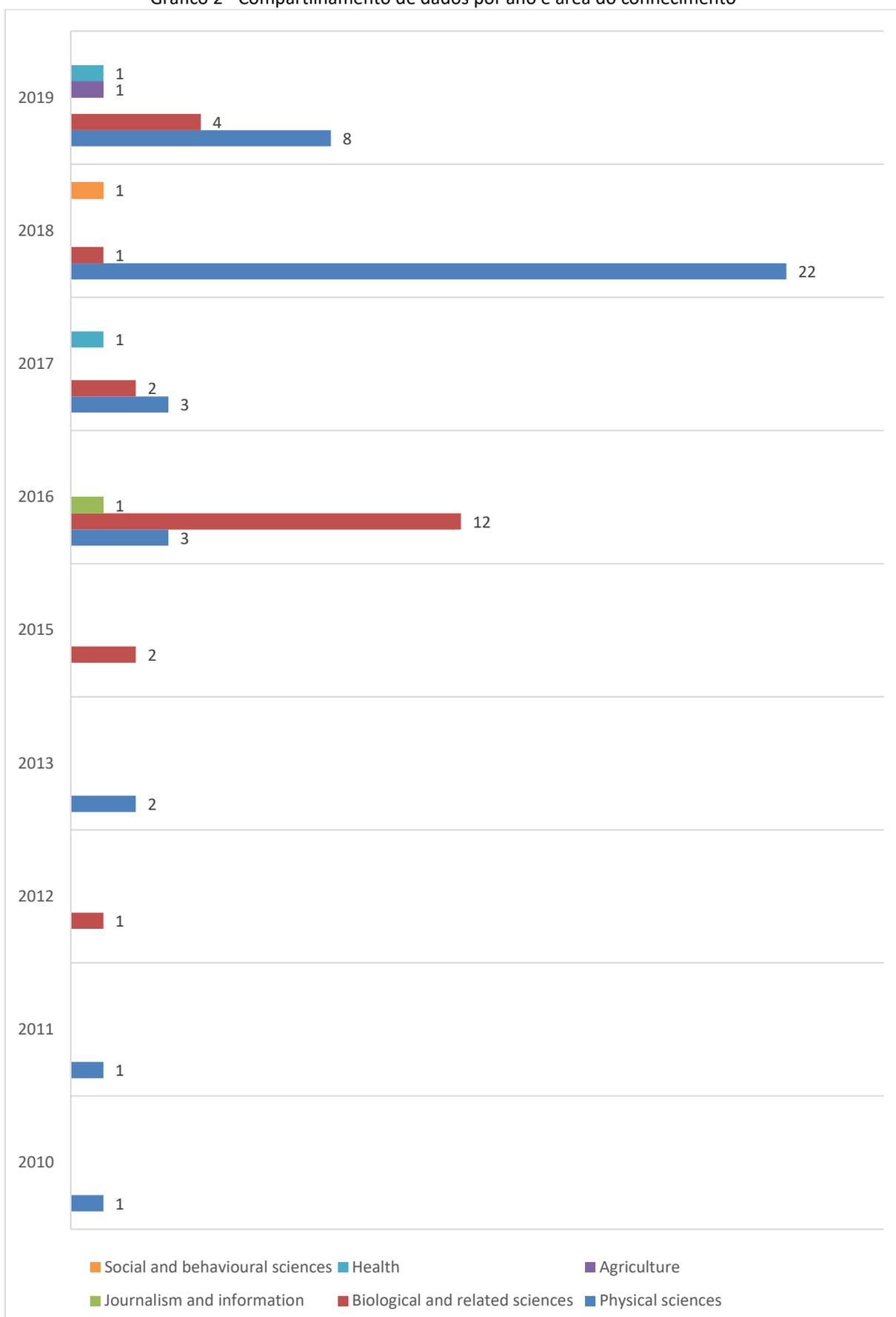


Fonte: Elaboração própria (2019).

A expressividade de compartilhamento de dados nas disciplinas *Physical sciences* (Ciências físicas) pode estar associada à presença do repositório *Cambridge Crystallographic Data Centre (CCDC)*, que representa 58% da amostra. Já o CCDC é um dos líderes mundiais em dados de química estrutural, software e conhecimento para pesquisa e aplicação em ciências da vida (CCDC, 2019).

Na amostra estudada, além do CCDC, verifica-se a presença de outros cinco repositórios de dados: *Dryad Digital Repository*, *Figshare*, *dat@OSU*, *SEANOE* e *University of Illinois at Urbana-Champaign*. Importa ressaltar que, de acordo com Banzi *et al.* (2019), os repositórios como o *Dryad* e *Figshare* concentram-se em serviços para compartilhamento de dados que suportam as descobertas relatadas em artigos de periódicos e outras publicações acadêmicas, além de serem multidisciplinares. Por exemplo, no *Dryad*, todo o material está associado a uma publicação acadêmica. Da mesma forma, o *Figshare* é um sistema de repositório disponível como um serviço de hospedagem na nuvem ou ao armazenamento local, o qual fornece serviços para armazenar e tornar objetos de dados visíveis vinculados aos dados publicados de vários editores científicos, como PLoS, F1000, Nature e Wiley (BANZI *et al.*, 2019). Destaca-se que quatro repositórios dos seis identificados, são repositórios multidisciplinares, isto é, concentram dados de diferentes áreas do conhecimento.

Gráfico 2 - Compartilhamento de dados por ano e área do conhecimento



Fonte: Elaboração própria (2019).

Observa-se que a área *Physical sciences* possui uma média de quatro compartilhamentos de dados por ano. Em 2018, há a expressão maior de compartilhamento de dados dessa área, com 22 registros vinculados a periódicos. Outro ponto a ser observado é a área *Biological and related sciences*, que no ano de 2016 apresentou 12 registros de dados vinculados a periódicos. Nesse mesmo ano ocorreu a publicação do trabalho de Wilkinson *et al.* (2016), intitulado de *Princípios orientadores da FAIR para gerenciamento e administração de dados científicos*, o qual é considerado a primeira publicação formal dos Princípios FAIR, porque inclui a lógica por trás deles e algumas implementações exemplares na comunidade. Outro marco é o início do desenvolvimento da política de compartilhamento de dados pela editora *Springer Nature*.

A quantidade de compartilhamentos de dados das áreas *Physical sciences* (Ciências físicas) e *Biological and related sciences* (Biologia e áreas correlatas) em relação às outras disciplinas é maior, pois a origem dos dados e a cultura acadêmica de como os pesquisadores comunicam seus resultados de pesquisa influenciam no seu compartilhamento. Como relatado na revisão de literatura deste trabalho, os repositórios de dados genéticos foram uma das iniciativas pioneiras internacionais, que mobilizaram pesquisadores do mundo inteiro. Quanto à natureza dos dados, os que não identificam pessoas e foram disponibilizados sem necessidade de análise são mais fáceis de compartilhar, ao contrário de dados das áreas *Health* (Saúde) e *Social and behavioural sciences* (Ciências sociais e comportamentais), que muitas vezes precisam identificar os indivíduos dos estudos, esbarrando em questões éticas.

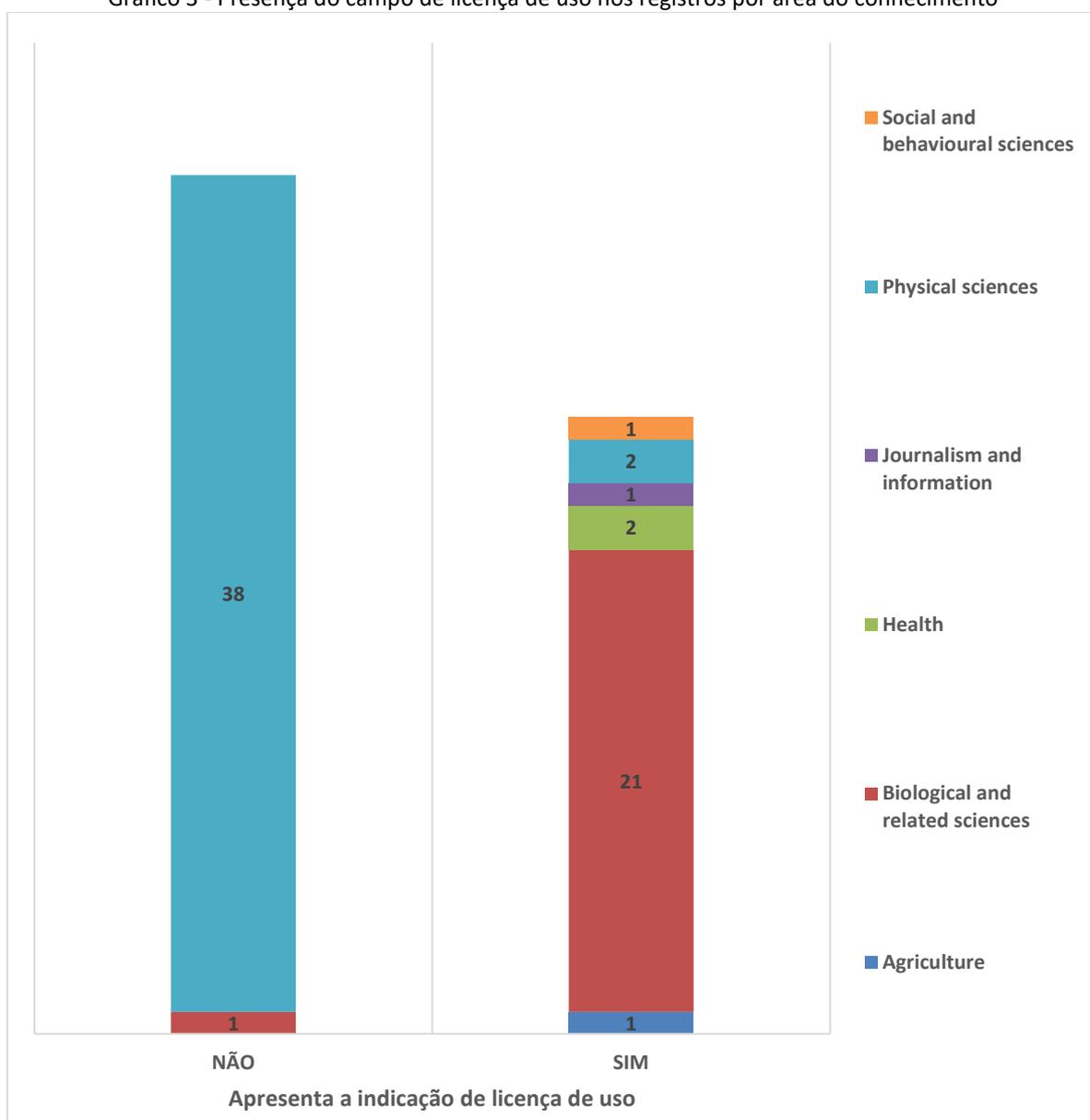
4.2 Análise do nível de descrição dos registros de conjunto de dados

Na amostra, identifica-se que 46% dos registros nos repositórios de dados estão em conformidade com os princípios FAIR elencados na Seção 3.2.3. Isto ocorre porque todos os registros possuem o campo que indica a versão ou a data de depósito e/ou alteração e, no mais, apresentam o campo “Como citar”. Com relação a esse último campo, observa-se que ele atende ao princípio R1.2, indicando como o autor gostaria de ser reconhecido, isto é, como citar o registro. Outro princípio atendido por esse campo é o F3, que explicita que os identificadores persistentes devem estar visíveis para os pesquisadores e, também, para o princípio R1.3.

Em relação à indicação de licença de uso, 58% dos registros não mencionam o tipo de licença (Gráfico 3). Aqueles que a apresentam, em sua maioria, exibem licenças *Creative Commons* do tipo mais aberto, isto é, a Marca de Domínio Público, a qual indica que está

livre de direitos autorais e em domínio público. Já a licença Atribuição CC BY permite distribuição, remix e adaptação, mesmo para fins comerciais, desde que lhe atribuam crédito pela criação original. Como afirma Guanaes (2020), o licenciamento para (re)utilização de dados de pesquisa ganha importância e precisa oferecer termos de uso claramente definidos, cabendo ao proprietário desses direitos marcar os dados com as permissões associadas. Nesse contexto, o autor relata que, entre outros tipos de licenciamento, as licenças *Creative Commons* têm sido adotadas por repositórios de dados, o que é uma forma de se alinhar às recomendações apontadas pelo FAIR no intuito de promover o compartilhamento e o reúso de obras criativas, educacionais e científicas.

Gráfico 3 - Presença do campo de licença de uso nos registros por área do conhecimento

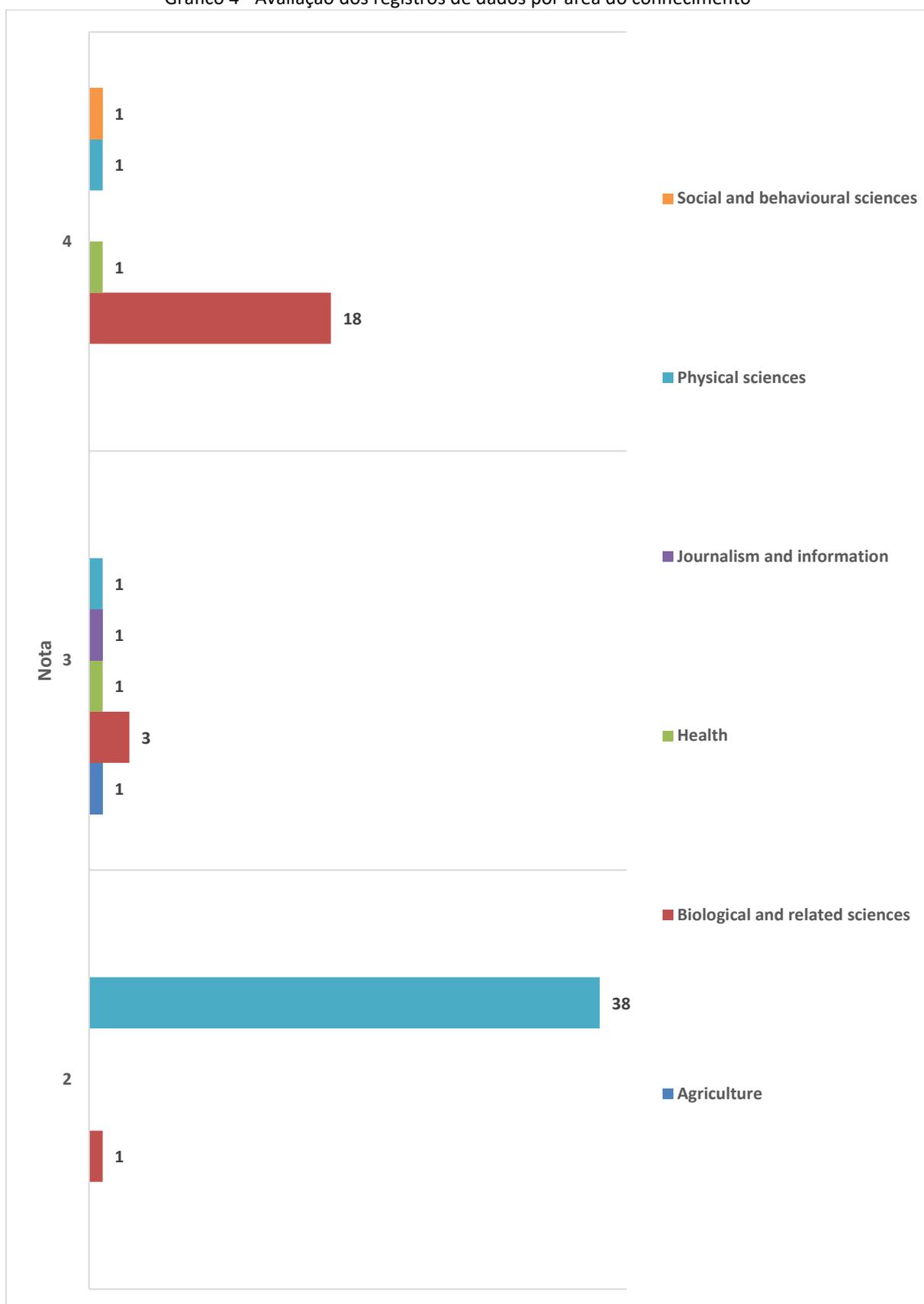


Fonte: Elaboração própria (2019).

Não foi possível averiguar o uso de vocabulário que seguem os FAIR, mas observou-se que nos quatro repositórios identificados há a presença de palavras-chave, as quais não têm qualquer relação com os princípios FAIR. Os autores Rautenberg e Souza (2020) defendem que a reutilização de vocabulários padronizados contribui ao reuso, à processabilidade, à compreensão, à confiança e à interoperabilidade de dados abertos, facilitando a inclusão de recursos informacionais da editoração científica e a expansão da Web de Dados. Por sua vez, Santarem Segundo (2020) concorda com tal ideia, ao afirmar que usar vocabulários reconhecidos internacionalmente é um dos pontos cruciais para que os dados possam ser coletados e reutilizados em aplicações diversas.

O último critério analisado foi a presença de um campo específico para descrever como os dados foram coletados, ou como e onde foram gerados. Da amostra, 69% dos registros não apresentam esse campo específico, porém observou-se que o campo destinado à descrição foi utilizado pelos autores para apresentar tais informações. Das áreas de conhecimento, a *Biological and related sciences* possui 82% dos registros que utilizam um campo específico para descrever essa informação sobre a coleta e geração dos dados – ao contrário da disciplina *Physical science*, que apresentou o campo em apenas 3% dos registros. A ausência de tal campo nessa área do conhecimento pode ser explicada pelo fato de que, na amostra, a maioria dos registros de conjuntos de dados advém do repositório CCDC, um repositório específico, ou porque a área da pesquisa determina uma metodologia própria para a comunicação dos dados.

Gráfico 4 - Avaliação dos registros de dados por área do conhecimento



Fonte: Elaboração própria (2019).

A partir da análise dos registros de conjunto de dados constata-se que há diferenças acerca do nível de descrição dos registros entre as áreas do conhecimento (Gráfico 4). A área do conhecimento que mais atendeu aos critérios foi a *Biological and related sciences* (85%). As outras áreas apresentaram apenas um registro com nota 4: o *Health, Physical sciences* e *Social and behavioural sciences*. Das cinco áreas identificadas, duas não alcançaram a nota 4, a saber: *Agriculture* e *Journalism and information*. Os dados analisados sugerem que a descrição dos registros de conjunto de dados está relacionada com as políticas dos repositórios de dados, que, por sua vez, são influenciadas pela cultura acadêmica de cada disciplina, no caso dos repositórios específicos de uma área.

4.3 Avaliar as políticas de depósito de dados de pesquisa de periódicos com o intuito de verificar a relação entre as políticas e as áreas do conhecimento

Para análise das políticas de compartilhamento de dados foram utilizadas duas amostras. Uma amostra com os 16 títulos de periódicos, para análise das políticas em si, isto é, a ficha de avaliação, descrita nos procedimentos metodológicos. Porém, quando foi realizada análise levando em consideração a área de conhecimento, utilizou-se a amostra com 18 revistas, visto que a *Royal Society Open Science* é multidisciplinar.

O compartilhamento de dados científicos é cada vez mais incentivado ou obrigado por financiadores e editores científicos (MENNES *et al.*, 2013). Sob essa perspectiva, na amostra 63% das revistas de periódicos apresentaram políticas de compartilhamento de dados. Entende-se que essas políticas estão localizadas na seção das políticas editoriais do periódico. Segundo Shintaku e Ferreira Junior (2020), uma política editorial, no sentido da editoração científica, pode ser entendida com um conjunto de diretrizes que orientam as atividades dos envolvidos na disseminação da informação científica.

Observa-se que, dos periódicos sem políticas de compartilhamento de dados (37%), apenas duas revistas (ISSN 0010-0870 e 0101-2061) não fazem qualquer menção ao compartilhamento de dados científicos. Ambas apresentam um artigo cada (Quadro 22), de modo que esse resultado pode estar associado à falta de incentivo da revista em relação à abertura dos dados de pesquisa. Entre as revistas que possuem mais de um artigo e que compartilham dados, o periódico *Chemical Science* (ISSN 2041-6539) é o que tem maior número: possui 35 registros de conjuntos de dados (Quadro 22), e a revista possui a política que encoraja a disponibilização de dados científicos.

Quadro 19 - Quantidade de compartilhamento de dados por periódico

	NOME DO PERIÓDICO	ISSN	QTD. DE REGISTROS	ÁREA DO CONHECIMENTO
1	Chemical Science	2041-6539	35	Physical sciences
2	Royal Society Open Science	2054-5703	11	Biological and related sciences Physical sciences Social and behavioural sciences
3	ZooKeys	1313-2989	4	Biological and related sciences
4	Ecology and Evolution	2045-7758	3	Biological and related sciences
5	Zeitschrift für Kristallographie - New Crystal Structures	1433-7266	2	Physical sciences
6	Molecular Systems Biology	1744-4292	2	Biological and related sciences
7	College and Research Libraries	0010-0870	1	Journalism and Information
8	Food Science and Technology	0101-2061	1	Agriculture
9	Biodiversity Data Journal	1314-2836	1	Biological and related sciences
10	Croatica Chemica Acta	1334-417X	1	Physical sciences
11	Kidney & Blood Pressure Research	1423-0143	1	Health
12	Bioinorganic Chemistry and Applications	1687-479X	1	Biological and related sciences
13	Clinical Epigenetics	1868-7083	1	Health
14	Rice	1939-8433	1	Biological and related sciences
15	BioTechniques	1940-9818	1	Biological and related sciences
16	Earth and Space Science	2333-5084	1	Physical sciences

Fonte: Elaboração da autora (2019).

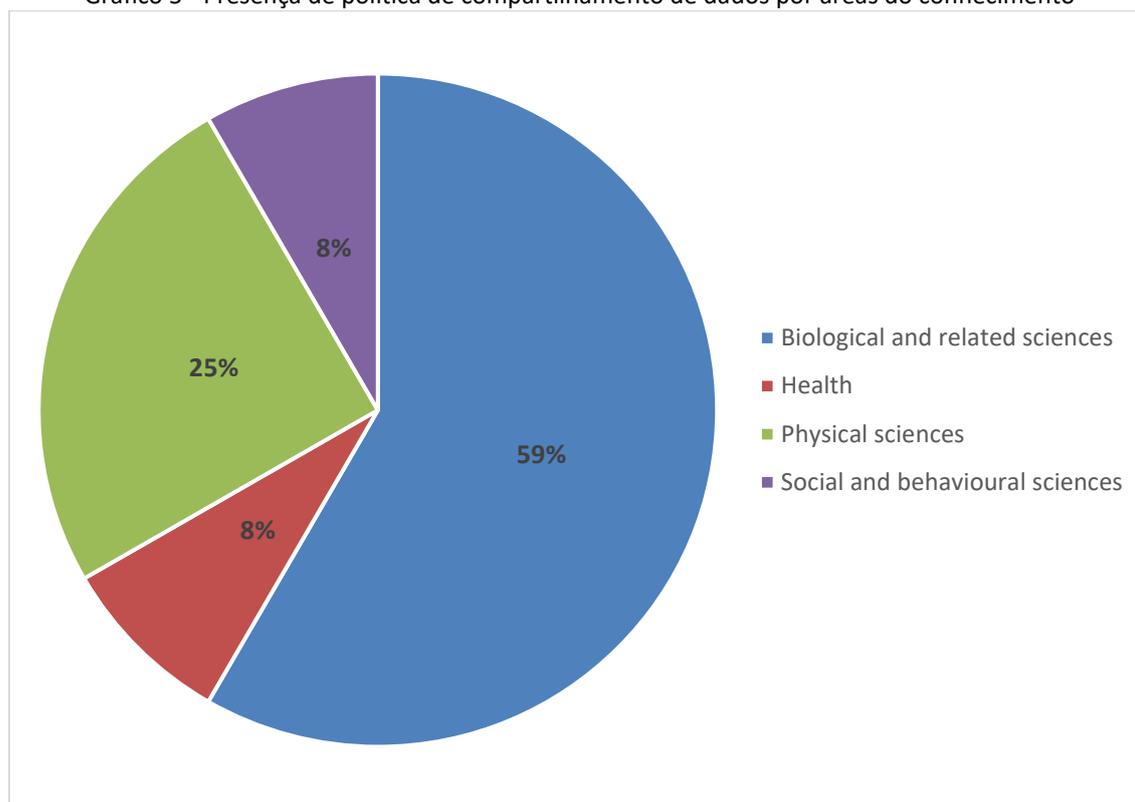
Dos 10 periódicos que possuem políticas de compartilhamento de dados, 50% obrigam o compartilhamento de dados para algum tipo de manuscrito. Nos casos das revistas que apresentam políticas de compartilhamento, a publicação dos dados pode ser realizada por meio de submissão dos conjuntos de dados a repositórios de dados ou inserida como material suplementar durante o processo de submissão do artigo à revista. No caso da última opção, em alguns dos exemplos analisados, a revista possui parceria com repositório de dados Dryad,¹⁴ uma organização que tem parcerias com várias casas editoriais e periódicos científicos importantes, os quais adotam para suas publicações acadêmicas uma Política Conjunta de Arquivamento de Dados mais conhecida pela sigla em inglês JDAP, que significa *Joint Data Archiving Policy*¹⁵ (SALES; SAYÃO, 2020). Após a publicação do artigo, esses conjuntos de dados são depositados, automaticamente, no repositório parceiro, recebendo um identificador persistente e o link entre o dado e a publicação/artigo científico. Para os periódicos que cobram pela publicação esse serviço de depósito em repositório também é cobrado.

Em relação à área do conhecimento, observa-se que as áreas da *Agriculture* e do *Journalism and information* não apresentam políticas de compartilhamento de dados (Gráfico 5). Em contrapartida, a área de *Biological and related sciences* representa mais da metade de periódicos com políticas de compartilhamento. Tal fato sugere a relação da área do conhecimento e a presença de política de compartilhamento, uma vez que – como apontado na literatura desta pesquisa – a *Biological and related sciences* possui um histórico de iniciativa e discussões no compartilhamentos de dados.

¹⁴ Disponível em: <https://datadryad.org/stash>.

¹⁵ Disponível em: [http://wiki.datadryad.org/Joint_Data_Archiving_Policy_\(JDAP\)](http://wiki.datadryad.org/Joint_Data_Archiving_Policy_(JDAP)).

Gráfico 5 - Presença de política de compartilhamento de dados por áreas do conhecimento



Fonte: Elaboração da autora (2019).

Da amostra estudada, 70% dos periódicos recomendam o depósito dos dados em repositório de dados. Para Sales e Sayão (2020), a criação de repositórios de dados multidisciplinares é a maneira mais adequada para viabilizar a interconexão de dados de pesquisa e publicações acadêmicas. Isto ocorre porque, como afirmam os autores, “a publicação de dados é parte imprescindível do processo de comunicação científica”. Com relação às áreas do conhecimento, observa-se que esses periódicos são das áreas *Biological and related sciences*, *Health* e *Physical sciences*, das 18 revistas analisadas – lembrando que a *Royal Society Open Science* foi contabilizada três vezes por causa do seu caráter multidisciplinar –, dez indicavam repositórios de dados. Desse total, seis eram da área de *Biological and related science*. Levando em consideração que uma das primeiras iniciativas de criação de banco de dados científicos foi o Genbank, a quantidade de políticas de compartilhamento de dados na área de *Biological and related science* encontrada no estudo pode ser devida ao histórico do compartilhamento de dados científicos, pois são alguns anos de construção de políticas e diretrizes para os pesquisadores da área. Com relação aos repositórios de dados, a amostra analisada identificou-se 6 repositórios de dados, sendo que 4 repositórios são considerados como repositórios multidisciplinares. Para Sales e Sayão (2020) a criação de repositórios de dados multidisciplinares é uma das maneiras para

viabilizar a interconexão de dados de pesquisa e publicações acadêmicas. Uma vez que, como afirmam os autores, “a publicação de dados é parte imprescindível do processo de comunicação científica”

Durante a verificação das políticas de compartilhamento de dados notam-se algumas diferenças e semelhanças entre os periódicos. Com esse mesmo objetivo foi acrescida à ficha de verificação uma coluna para observações a respeito de cada política. Assim, destacam-se os seguintes pontos das revistas:

ZooKeys (área do conhecimento: *Biological and related sciences*) possui uma política que encoraja o compartilhamento de dados, recomenda o uso de repositório de dados ao invés de arquivamento como material suplementar, indica aos autores o estudo realizado por Penev *et al.* (2017), o qual aponta as estratégias e diretrizes para publicação acadêmica de dados sobre biodiversidade, i.e., o conceito de conjunto de dados, como publicar e citar dados, onde publicar os dados, licenças etc. Além de recomendações para os autores, há orientações para os revisores e editores no estudo dos autores. Por fim, o periódico apresenta uma lista de verificação e recomendações sobre qualidade de dados no intuito de maximizar a reutilização dos dados para os autores. Nessa lista abordam-se algumas características necessárias, como codificação, uso de caracteres válidos e preenchimento de campos.

Biodiversity Data Journal (área do conhecimento: *Biological and related sciences*) possui uma política que obriga o compartilhamento de dados, oferta aos autores as opções de importar arquivos de dados no texto, arquivar como material suplementar, depositar em repositórios internacionais estabelecidos ou envio de arquivo no formato XML para garantir a colheita da máquina. Este periódico também indica aos autores, o estudo realizado por Penev *et al.* (2017), o qual aponta as estratégias e diretrizes para publicação acadêmica de dados sobre biodiversidade. Como este periódico é da editora Pensoft, a mesma editora da revista Zookeys, também apresenta uma lista de verificação e recomendações sobre qualidade de dados no intuito de maximizar a reutilização dos dados para os autores. Além disso, este periódico possui parceria com o repositório *Dryad Digital Repository*, mas o serviço é cobrado.

Molecular Systems Biology (área do conhecimento: *Biological and related sciences*) apresenta uma seção sobre a disponibilização dos dados, fornece uma lista de repositórios de dados e, também, a opção de incluir como material suplementar. Nesse periódico não é permitido citar dados que não podem ser mostrados, isto é, todo e qualquer dado utilizado no estudo deve ser citado com a indicação do identificador único. Por ser obrigatória a

disponibilização dos dados, a política ressalta a importância do cumprimento, dos padrões éticos e da integridade de dados. Já no caso de dados que possam ter relação com biossegurança ou conflitos de natureza acadêmica ou comercial, os editores precisam ser comunicados.

Clinical Epigenetics (área do conhecimento: *Health*) por ser uma revista do grupo BMC (faz parte da editora Springer Nature) adota a política desse grupo. A política, de forma geral, encoraja o compartilhamento de dados para qualquer uma das revistas pertencentes ao grupo BMC. O compartilhamento de dados segue as orientações do *Force 11 Data Citation Principles* e apresenta uma lista de repositórios de dados categorizada por disciplinas da área da saúde. Não há obrigatoriedade em compartilhar, isto é, a submissão, – e, conseqüentemente, a aprovação – do artigo não será negada caso não haja o compartilhamento e/ou citação dos dados, porém, é necessário que se justifiquem os motivos, de modo que a justificativa será avaliada. Tal prática, em certo ponto, é uma forma de exigir dos pesquisadores a disponibilização dos dados, pois a publicação do trabalho na revista está condicionada a esse requisito.

Rice (área do conhecimento: *Biological and related sciences*) é uma revista da editora Springer Nature que faz parte do grupo de revistas de acesso aberto dessa editora, intitulado *SpringerOpen*. A política de compartilhamento assemelha-se à política da revista **Clinical Epigenetics**, pois são da mesma editora. O que difere é o fato de que, como o acesso ao periódico **Rice** é direto pela página da *SpringerOpen*, a lista de repositório é mais extensa, englobando várias áreas do conhecimento, como química, genética, biologia etc.

BioTechniques (área do conhecimento: *Physical sciences*) não apresenta uma política detalhada de compartilhamento, isto é, uma seção exclusiva para tratar sobre a disponibilização de dados como os periódicos mencionados acima. Nas políticas editoriais há uma seção sobre depósito de dados, porém não há indicação de repositórios. Na seção destinada às orientações para os autores explana-se sobre a importância de depositar os dados em repositórios ou banco de dados apropriados, respeitar as questões éticas e citar os dados advindos de outras fontes. Além disso, o periódico ressalta que, para alguns tipos estudos, como o estatístico, é imprescindível a apresentação dos dados brutos. Já em relação aos softwares, os padrões internacionais devem ser respeitados.

Chemical Science (área do conhecimento: *Physical sciences*) é uma revista da *Royal Society of Chemistry* e, portanto, a política editorial, inclusive a política de compartilhamento de dados, é única para todas as revistas administradas pela entidade. O compartilhamento de

dados não é obrigatório, mas é encorajado. Há indicação de repositórios para o depósito de dados e, também, uma descrição de como os dados devem ser preparados de forma geral e de acordo com sua natureza.

A revista **Ecology and Evolution** (área do conhecimento: *Biological and related sciences*) apresenta uma política que obriga o compartilhamento de dados, a fim de elucidar que os dados, utilizados no estudo ou gerados por ele, devem estar em repositórios ou no banco de dados de acesso aberto. Para os dados gerados/coletados pela pesquisa é necessário que o autor assine e/ou concorde com a Declaração de Acessibilidade de Dados (*Data Accessibility Statement*), de modo que esse documento deve ser enviado durante o processo de submissão do artigo. A lista de repositório não é extensa, pois apresenta apenas repositórios gerais, como Dryad e Figshare, e alguns temáticos, como o GenBank.

O periódico **Royal Society Open Science** (áreas do conhecimento: *Biological and related sciences, Social and behavioural sciences e Physical sciences*) apresenta uma política na qual exige-se que dados e informações suplementares, incluindo código-fonte e outros materiais de pesquisa, sejam disponibilizados no momento da submissão. Para a revista, essa prática se alinha às suas políticas que visam promover maior abertura na pesquisa científica e permitir, além de incentivar, outros pesquisadores a executar repetições completas dos estudos publicados. Como forma de facilitar o cumprimento dessa política, o sistema de submissão é integrado ao repositório de dados Dryad, permitindo que o autor realize o *upload* do material no repositório e informe aos editores e revisores apenas o link compartilhável na Declaração de Acessibilidade dos Dados, só que a utilização desse repositório pode ser cobrada. Outra opção é enviar os dados como material suplementar, de maneira que, após a aceitação dos manuscritos, eles serão depositados no portal Figshare da Royal Society gratuitamente. Além dessas alternativas, os autores podem enviar seus dados para outros repositórios de dados acessíveis ao público (por exemplo, repositórios de universidades, Zenodo, Open Science Framework etc.).

A última política analisada pertence ao periódico **Earth and Space Science** (área do conhecimento: *Physical sciences*) e estabelece que os dados devem ser depositados em um repositório confiável, que pratique os princípios FAIR. Dispõe aos autores orientações de como citar, como selecionar repositórios e como modelar os dados. Além disso, oferece modelos de elaboração da Declaração de Acessibilidade dos Dados.

Dentre as revistas que possuem políticas de compartilhamento de dados, nota-se que, em sua maioria, tal política está ligada à política editorial, isto é, trata-se de uma seção dentro

da política da revista. Isto pode facilitar a busca por informações dos autores, pois em único documentou e/ou página na internet encontram-se as diretrizes éticas e de integridade em pesquisa. Ademais, observa-se nas políticas analisadas a preocupação com as questões éticas ligadas aos dados, isto é, não divulgar informações pessoais (anonimizar os dados), apresentar declaração de que a pesquisa foi aceita pelo comitê de ética, seguir recomendações internacionais etc. Tal preocupação alinha-se ao lema da Ciência Aberta no que tange à gestão de dados de pesquisa, a qual diz: “Tão aberto quanto possível, tão fechado quanto necessário” (MALAPELA, 2016). Os autores Chauvette, Schick-Makaroff e Molzahn (2019) argumentam que, para assegurar a disponibilização dos dados de pesquisa, as questões epistemológicas, metodológicas e éticas para abertura dos dados devem estar amparadas, com vistas a permitir essa segurança no compartilhamento.

5 CONSIDERAÇÕES FINAIS

O compartilhamento de dados ainda não é unanimidade entre os pesquisadores devido às diferenças entre as disciplinas geográficas (marcação de políticas governamentais, agências de fomento etc.), econômicas e sociais. Porém, para atingir os objetivos da ciência aberta é preciso comunicar os dados de pesquisa. Além disso, a abertura dos dados de pesquisa justifica-se pelos princípios científicos de transparência, colaboração e refutabilidade, contribuindo para o desenvolvimento científico.

Diante do cenário de compartilhamento de dados abordado neste estudo, observa-se que a cultura de comunicação de dados difere entre as áreas do conhecimento. As comunidades científicas de cada disciplina passaram a desenvolver seus sistemas de repositórios de dados com o objetivo de atender às suas necessidades e tradições de compartilhamento.

Algumas áreas do conhecimento já estão habituadas, em certo nível, a compartilhar os dados de pesquisa, como a área *Physical sciences*, que neste estudo possui uma média de compartilhamento maior do que as outras áreas. A justificativa para essa quantidade maior de registros de conjuntos de dados pode estar associada às características, por exemplo, epistemológicas, metodológicas e éticas existentes, pois tais características influenciam diretamente no compartilhamento de dados de pesquisa. Contudo, o compartilhamento de dados ocorre também nas outras áreas do conhecimento, como verificado neste estudo.

Nos compartilhamentos analisados observou-se que o reuso desses dados não é tão expressivo quanto a disponibilização deles. Fato é que apenas três registros deste estudo indicam a probabilidade de que os dados referenciados nos artigos de periódicos não foram coletados pelos mesmos autores, e sim reutilizados por eles. A verificação desses indícios ocorre por meio da presença do metadado *relType*, que indica se o registro de conjunto de dados é suplemento (*IsSupplementTo*) ou citado (*IsCitedBy*) do/no artigo científico. A presença desse campo é amparada pelos os princípios FAIR, ou seja, é um metadado necessário para tornar o dado mais “FAIR”.

Os registros de conjuntos de dados analisados neste estudo, em sua maioria, seguem alguns dos princípios FAIR. Principalmente porque a elaboração do *script* e onde foi realizada a coleta de dados (no DataCite) já atende a algumas das orientações FAIR, como: atribuir identificador persistente aos registros de conjuntos de dados, indexar os metadados, utilizar protocolos internacionais de comunicação e os metadados que indicam a relação entre

o registro de conjunto de dados e os artigos de periódicos. Contudo, ressalta-se a ausência do uso de vocabulários controlados que atendam aos princípios FAIR, de modo que a análise não foi possível porque não estava explícita, nos campos de vocabulário/palavras-chave, a indicação de qualquer vocabulário, bem como nas políticas de depósitos dos repositórios também não foi encontrada qualquer indicação para uso de vocabulário específico. O princípio I2 ressalta que o vocabulário controlado usado para descrever conjuntos de dados precisa ser documentado e disponibilizado usando identificadores persistentes e únicos. Esta documentação precisa ser facilmente encontrada e estar acessível a qualquer pessoa que use o conjunto de dados a fim de permitir que esses dados sejam interoperáveis.

Nos registros de dados, a presença do campo **Como citar** é unânime, pois todos os registros analisados possuem este campo. Tal alternativa, em certa medida, facilita o reuso dos dados por outros pesquisadores, pois indica como o registro deve ser referenciado. Outro campo que foi constatado em todos os registros é o que indica a versão ou data de depósito e/ou alteração. A maioria dos registros que não apresentaram esse campo foi da área da *Physical Science*, que em sua maioria eram de um repositório de dados, o *Cambridge Crystallographic Data Centre*. Isto pode levar a questionamentos acerca da política e do esquema de metadados utilizados por esse repositório. Outro ponto observado nesse mesmo repositório é a falta do campo indicando a licença de uso. Para os princípios FAIR, a indicação da licença de uso é observada no princípio R1.1, que trata de interoperabilidade legal. Além disso, menos da metade dos registros apresentam esse campo.

Com relação ao repositório *Cambridge Crystallographic Data Centre*, outro ponto observado foi a falta do metadado que informa como os dados foram obtidos, as condições, a localização geográfica etc. A falta de alguns metadados nos registros e o alto número de compartilhamento de dados podem estar relacionados. Como o depósito de dados é realizado pelo pesquisador, um processo de submissão simplificado torna o depósito dos dados mais simples para os pesquisadores, por ser mais rápido. Porém, outro fator que pode ser considerado é a obrigatoriedade do compartilhamento de dados por parte dos periódicos científicos, ou seja, ser requisito para submissão de artigos científicos nas revistas.

Com relação às políticas de depósito de dados de pesquisa, verificou-se que dois dos três periódicos da área *Physical Science* obrigam o compartilhamento de dados como requisito para submissão de artigos, fato esse que pode explicar o porquê de neste estudo haver mais compartilhamentos nessa disciplina. De maneira geral, as revistas vêm trabalhando em suas políticas de dados, tanto é que mais da metade das revistas analisadas

possuem políticas de dados. Essas políticas majoritariamente indicam quais são os repositórios de dados preferenciais para o depósito dos dados de pesquisa.

Durante a coleta de dados, a interface API disponibilizada pelo DataCite não permitiu a coleta total dos dados, isto é, acredita-se que o resultado retornado pelo *script* não corresponde ao valor total da base do DataCite. Pois, ao se realizar a consulta na base, o resultado obtido foi de aproximadamente 6 milhões de registros do tipo *dataset*. Contudo, só foram recuperados 33.245 registros, os quais foram extraídos de forma aleatória, porque quando não aplicado o filtro *random* (aleatório) a recuperação não era possível. A dificuldade de acesso à base pode ser considerada uma barreira para os estudos sobre compartilhamento de dados, uma vez que muitos repositórios estão associados ao DataCite.

Contribuição do estudo

O estudo apresenta um panorama do compartilhamento de dados por algumas áreas do conhecimento no período de dez anos, com o objetivo de identificar periódicos científicos cadastrados no DOAJ, suas políticas editoriais sobre dados e os campos utilizados para descrição dos registros de conjuntos de dados nos repositórios de dados.

No campo da CI esta pesquisa apresenta-se como uma contribuição para as discussões da comunicação científica. Nesse sentido, o compartilhamento dos dados de pesquisa se faz presente no fazer ciência, assim como é parte essencial para a consolidação dos objetivos da Ciência Aberta. Portanto, este estudo preenche as lacunas de conhecimento do campo, ao mesmo tempo que enriquece a discussão.

Sugestão de estudos

Neste estudo, evidencia-se o uso do metadado *relType* nos repositórios de dados como formas de indicar o compartilhamento dos dados. No entanto, o tema não foi aprofundado por não se tratar do objetivo desta pesquisa. Portanto, considera-se que analisar a reutilização dos dados depositados em repositórios por meio desse metadado é relevante para estudos futuros, pois um dos intuitos de compartilhar dados é o reúso deles por outros pesquisadores.

As diferenças entre as disciplinas foram o objeto central deste estudo, pois as culturas acadêmicas de comunicar cientificamente tornam-se importantes os estudos sobre o fenômeno do compartilhamento de dados. Entretanto, os repositórios específicos de cada área do conhecimento não foram abordados a fundo e, por isso, recomenda-se que estudos futuros

considerem analisar as políticas de depósito dos dados dos repositórios de dados temáticos para a explicação dos fenômenos acerca da comunicação dos dados.

REFERÊNCIAS

- ABBOTT, D. What is Digital Curation? **DCC Briefing Papers: Introduction to Curation**, 2008. Disponível em: <https://www.dcc.ac.uk/guidance/briefing-papers/introduction-curation/what-digital-curation>. Acesso em: 9 set. 2020.
- ALBAGLI, S. Ciência aberta em questão. *In*: ALBAGLI, S.; MACIEL, M. L.; ABDO, A. H. (orgs.). **Ciência aberta, questões abertas**. Brasília: IBICT, 2015. p. 9–25. Disponível em: <http://livroaberto.ibict.br/handle/1/1060>. Acesso em: 9 set. 2020.
- ALBAGLI, S.; CLINIO, A.; RAYCHTOCK, S. Ciência Aberta: correntes interpretativas e tipos de ação. **Liinc em Revista**, v. 10, n. 2, 5 dez. 2014. DOI 10.18617/liinc.v10i2.749. Disponível em: <http://revista.ibict.br/liinc/article/view/3593>. Acesso em: 9 set. 2020.
- ALLARD, S.; MACK, T. R.; FELTNER-REICHERT, M. The librarian’s role in institutional repositories: A content analysis of the literature. **Reference Services Review**, v. 33, n. 3, p. 325–336, set. 2005. DOI 10.1108/00907320510611357. Disponível em: <https://www.emerald.com/insight/content/doi/10.1108/00907320510611357/full/html>. Acesso em: 9 set. 2020.
- ALLEN, R.; HARTLAND, D. **Fair In Practice - Jisc Report On The Findable Accessible Interoperable And Reuseable Data Principles**. [S. l.]: Zenodo, 21 maio 2018. DOI 10.5281/ZENODO.1245568. Disponível em: <https://zenodo.org/record/1245568>. Acesso em: 11 set. 2020.
- ANDREOLI-VERSBACH, P.; MUELLER-LANGER, F. Open access to data: An ideal professed but not practised. **Research Policy**, v. 43, n. 9, p. 1621–1633, nov. 2014. DOI 10.1016/j.respol.2014.04.008. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0048733314000699>. Acesso em: 9 set. 2020.
- ARAKAKI, F. A. **Metadados administrativos e a proveniência dos dados: modelo baseado na família PROV**. 2019. 139 f. Tese (Doutorado em Ciência da Informação) – Universidade Estadual Paulista “Júlio Mesquita Filho”, Marília, SP, 2019. Disponível em: <https://repositorio.unesp.br/handle/11449/180490>. Acesso em: 24 set. 2020.
- ARAKAKI, F. A.; ARAKAKI, A. C. S. Dados e metadados: reflexões conceituais. *In*: WIDAT 2019, 2019. **Anais** [...]. Brasília: UnB, 2019. p. 60–65. Disponível em: <http://wizat2019.fci.unb.br/index.php/anais-wizat-2019>. Acesso em: 24 set. 2020.
- AULER, M. da S. **A produção do conhecimento publicada no periódico história, ciências, saúde-manguinhos: um olhar sobre as citações**. 2019. 103 f. Dissertação (Mestrado em Informação e Comunicação em Saúde) – Instituto de Comunicação e Informação Científica e Tecnológica em Saúde, Fundação Oswaldo Cruz, Rio de Janeiro, 2019. Disponível em: <https://www.arca.fiocruz.br/handle/icict/34424>. Acesso em: 9 set. 2020.

AVENTURIER, P.; ALENCAR, M. C. de. Os desafios de dados de pesquisa abertos. **Revista Eletrônica de Comunicação, Informação e Inovação em Saúde**, v. 10, n. 3, 30 set. 2016. DOI 10.29397/reciis.v10i3.1069. Disponível em: <https://www.reciis.icict.fiocruz.br/index.php/reciis/article/view/1069>. Acesso em: 9 set. 2020.

AZEVEDO, P. R. M. de. **Introdução à estatística**. 3. ed. Natal: EDUFERN, 2016. Disponível em: <https://repositorio.ufrn.br/jspui/handle/123456789/21298>. Acesso em: 16 set. 2020.

BANZI, R.; CANHAM, S.; KUCHINKE, W.; KRLEZA-JERIC, K.; DEMOTES-MAINARD, J.; OHMANN, C. Evaluation of repositories for sharing individual-participant data from clinical studies. **Trials**, v. 20, n. 1, p. 169, dez. 2019. DOI 10.1186/s13063-019-3253-3. Disponível em: <https://trialsjournal.biomedcentral.com/articles/10.1186/s13063-019-3253-3>. Acesso em: 9 set. 2020.

BENTANCOURT, S. M. P.; ROCHA, R. P. da. Metadados de qualidade e visibilidade na comunicação científica. **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação**, Florianópolis, p. 82-101, dez. 2012. ISSN 1518-2924. DOI: 10.5007/1518-2924.2012v17nesp2p82. Disponível em: <https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2012v17nesp2p82/23571>. Acesso em: 09 set. 2020.

BIOJONE, M. R. **Os periódicos científicos na comunicação da ciência**. São Paulo, Brasil: EDUC, FAPESP, 2003.

BJÖRK, B.-C. A model of scientific communication as a global distributed information system. **Information Research**, v. 12, n. 2, 2007. Disponível em: <http://informationr.net/ir/12-2/paper307>. Acesso em: 9 set. 2020.

BORGMAN, C. L. Research Data: Who Will Share What, with Whom, When, and Why? **SSRN Electronic Journal**, 2010. DOI 10.2139/ssrn.1714427. Disponível em: <http://www.ssrn.com/abstract=1714427>. Acesso em: 9 set. 2020.

BORGMAN, C. L. The conundrum of sharing research data. **Journal of the American Society for Information Science and Technology**, v. 63, n. 6, p. 1059–1078, jun. 2012. DOI 10.1002/asi.22634. Disponível em: <http://doi.wiley.com/10.1002/asi.22634>. Acesso em: 9 set. 2020.

BOULTON, G. Reinventing Open Science for the 21st Century. *In*: RODRIGUES, E.; SWAN, A.; BAPTISTA, A. A. (orgs.). **Uma década de acesso aberto na UMinho e no Mundo**. Lisboa: Universidade do Minho, 2013. p. 239–250. Disponível em: <http://hdl.handle.net/1822/26144>. Acesso em: 9 set. 2020.

BRASE, J. Making Data Citeable: DataCite. *In*: BARTLING, S.; FRIESIKE, S. (orgs.). **Opening Science**. Cham: Springer International Publishing, 2014. p. 327–329. DOI 10.1007/978-3-319-00026-8_26. Disponível em: http://link.springer.com/10.1007/978-3-319-00026-8_26. Acesso em: 25 set. 2020.

BRASE, J.; LAUTENSCHLAGER, M.; SENS, I. The Tenth Anniversary of Assigning DOI Names to Scientific Data and a Five Year History of DataCite. **D-Lib Magazine**, v.

21, n. 1/2, jan. 2015. DOI 10.1045/january2015-brase. Disponível em: <http://www.dlib.org/dlib/january15/brase/01brase.html>. Acesso em: 9 set. 2020.

BRASIL. Cartilha Técnica para Publicação de Dados Abertos no Brasil. 2019. **Portal Brasileiro de Dados Abertos**. Disponível em: <http://dados.gov.br/pagina/cartilha-publicacao-dados-abertos>. Acesso em: 6 set. 2020.

CAMBRIDGE CRYSTALLOGRAPHIC DATA CENTRE (CCDC). About us. 2019. **Cambridge Crystallographic Data Centre (CCDC)**. Disponível em: <https://www.ccdc.cam.ac.uk/theccdcprofile/>. Acesso em: 9 set. 2020.

CARVALHO, É. R. S. de; LEITE, F. C. L. Diferenças na produção, no compartilhamento e no (re)uso de dados de pesquisa: a percepção de pesquisadores de Química, Antropologia e Educação. **Em Questão**, v. 25, n. 3, p. 321–347, 7 ago. 2019. DOI 10.19132/1808-5245253.321-347. Disponível em: <https://seer.ufrgs.br/EmQuestao/article/view/89449>. Acesso em: 9 set. 2020.

CAVALCANTI, M. T.; SALES, L. F. Gestão de dados de pesquisa: um panorama da atuação da União Europeia. **BIBLOS**, v. 31, n. 1, p. 73–98, 5 ago. 2017. DOI 10.14295/biblos.v31i1.5789. Disponível em: <https://periodicos.furg.br/biblos/article/view/5789>. Acesso em: 9 set. 2020.

CHAUVETTE, A.; SCHICK-MAKAROFF, K.; MOLZAHN, A. E. Open Data in Qualitative Research. **International Journal of Qualitative Methods**, v. 18, p. 160940691882386, jan. 2019. DOI 10.1177/1609406918823863. Disponível em: <http://journals.sagepub.com/doi/10.1177/1609406918823863>. Acesso em: 9 set. 2020.

CHAVAN, V.; PENEV, L. The data paper: a mechanism to incentivize data publishing in biodiversity science. **BMC bioinformatics**, v. 12 Suppl 15, p. S2, 2011. <https://doi.org/10.1186/1471-2105-12-S15-S2>.

CHOUDHURY, G. S. The Virtual Observatory Meets the Library. **The Journal of Electronic Publishing**, v. 11, n. 1, 30 jan. 2008. DOI 10.3998/3336451.0011.111. Disponível em: <http://hdl.handle.net/2027/spo.3336451.0011.111>. Acesso em: 9 set. 2020.

COLLINS, S.; GENOVA, F.; HARROWER, N.; HODSON, S.; JONES, S.; LAAKSONEN, L.; MIETCHEN, D.; PETRAUSKAITÉ, R.; WITTENBURG, P. **Turning FAIR into reality: final report and action plan from the European Commission expert group on FAIR data**. LU: European Commission Expert Group on FAIR Data, 2018. Disponível em: <https://data.europa.eu/doi/10.2777/1524>. Acesso em: 11 set. 2020.

CORRÊA, Elizabeth Saad; BERTOCCHI, D. O papel do comunicador num cenário de curadoria algorítmica de informação. In: CORRÊA, Elisabeth Saad (org.). **Curadoria digital e o campo da comunicação**. São Paulo: ECA-USP, 2012. p. 22–39. Disponível em: <https://danielabertocchi.com/2012/10/09/o-algoritmo-curador-curadoria-digital-2012/>. Acesso em: 9 set. 2020.

COSTA, M. P. da. **Características e contribuições da via verde para o acesso aberto à informação científica na América Latina**. 2014. 226 f. Dissertação (Mestrado em Ciência da Informação) – Universidade de Brasília, Faculdade de Ciência da Informação, Programa de Pós-Graduação em Ciência da Informação, Brasília, 2014. Disponível em: <https://repositorio.unb.br/handle/10482/15687>. Acesso em: 9 set. 2020.

COSTA, M. P. da. **Fatores que influenciam a comunicação de dados de pesquisa sobre o vírus da zika, na perspectiva de pesquisadores.** 2017. 269 f. Tese (Doutorado em Ciência da Informação) – Universidade de Brasília, Faculdade de Ciência da Informação, Programa de Pós-Graduação em Ciência da Informação, Brasília, 2017. Disponível em: <https://repositorio.unb.br/handle/10482/23000>. Acesso em: 9 set. 2020.

COSTA, M. P.; LEITE, F. C. L. Princípios e recomendações basilares para a comunicação dos dados de pesquisa. **Em Questão**, v. 23, n. 1, p. 87, 22 dez. 2017. DOI 10.19132/1808-5245231.87-112. Disponível em: <http://seer.ufrgs.br/index.php/EmQuestao/article/view/65623>. Acesso em: 25 set. 2020.

COSTA, S. M. de S. **The impact of computer usage on scholarly communication amongst academic social scientists.** 1999. 318 f. Tese (Doutorado em Ciência da Informação) – Loughborough University, Londres, 1999. Disponível em: <https://repositorio.unb.br/handle/10482/2025>. Acesso em: 9 set. 2020.

COSTA, S. M. de S.; LEITE, F. C. L. Imbricações teóricas entre comunicação e gestão da informação e do conhecimento na Ciência da Informação. **Investigación Bibliotecológica: archivonomía, bibliotecología e información**, v. 32, n. 74, p. 225, 22 fev. 2018. DOI 10.22201/iibi.24488321xe.2018.74.57923. Disponível em: <http://rev-ib.unam.mx/ib/index.php/ib/article/view/57923>. Acesso em: 9 set. 2020.

COSTA, S. M. S.; MOREIRA, A. C. The diversity of trends, experiences and approaches in electronic publishing: evidences of a paradigm shift on communication. In: ELPUB CONFERENCE ON ELECTRONIC PUBLISHING, 7., 2003, Portugal. **Proceedings [...]** Portugal: Universidade do Minho, 2003. Disponível em: <https://repositorio.unb.br/handle/10482/1024>. Acesso em: 9 set. 2020.

CREATIVE COMMONS. CC0. 2020. **Creative Commons.** Disponível em: <https://creativecommons.org/share-your-work/public-domain/cc0/>. Acesso em: 18 set. 2020.

CRESWELL, J. W. **Projeto de pesquisa: métodos qualitativo, quantitativo e misto.** Porto Alegre: Sage, 2010.

CURTY, R. G.; AVENTURIER, P. O paradigma da publicação de dados e suas diferentes abordagens. In: ENANCIB 2017 : ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 2017., Accession Number: fdi:010071119source: IRD - Base Horizon / Pleins textes. **Anais [...]**. Marília: ENANCIB, 2017. Disponível em: <http://www.documentation.ird.fr/hor/fdi:010071119>. Acesso em: 9 set. 2020.

DALLMEIER-TIESSEN, S.; DARBY, R.; GITMANS, K.; LAMBERT, S.; MATTHEWS, B.; MELE, S.; SUHONEN, J.; WILSON, M. Enabling Sharing and Reuse of Scientific Data. **New Review of Information Networking**, v. 19, n. 1, p. 16–43, 2 jan. 2014. DOI 10.1080/13614576.2014.883936. Disponível em: <http://www.tandfonline.com/doi/abs/10.1080/13614576.2014.883936>. Acesso em: 9 set. 2020.

DATA CITE METADATA WORKING GROUP. **DataCite Metadata Schema Documentation for the Publication and Citation of Research Data v4.3.** , p. 73 pages,

2019. DOI 10.14454/7XQ3-ZF69. Disponível em: <https://schema.datacite.org/meta/kernel-4.3/>. Acesso em: 9 set. 2020.

DAVILA, V. H. L. Estatística Descritiva [Apresentação de slides]. 2007. Disponível em: <https://www.ime.unicamp.br/~hlauchos/estdescr1.pdf>. Acesso em: 16 set. 2020.

DAVIS, H. M.; VICKERY, J. N. Datasets, a Shift in the Currency of Scholarly Communication: Implications for Library Collections and Acquisitions. **Serials Review**, v. 33, n. 1, p. 26–32, mar. 2007. DOI 10.1080/00987913.2007.10765089. Disponível em: <http://www.tandfonline.com/doi/abs/10.1080/00987913.2007.10765089>. Acesso em: 9 set. 2020.

DELFANTI, A.; PITRELLI, N. Ciência aberta: revolução ou continuidade? *In*: ALBAGLI, S.; MACIEL, M. L.; ABDO, A. H. (orgs.). **Ciência aberta, questões abertas**. Brasília: IBICT, 2015. p. 59–78. Disponível em: <http://livroaberto.ibict.br/handle/1/1060>. Acesso em: 9 set. 2020.

DIGITAL CURATION CENTRE (DCC). What is digital curation? 2019. **DCC**. Disponível em: <https://www.dcc.ac.uk/about/digital-curation>. Acesso em: 9 set. 2020.

DIRECTORY OF OPEN ACCESS JOURNALS (DOAJ). About Directory of Open Access Journals. 2019. Disponível em: <https://doaj.org/about>. Acesso em: 15 set. 2020.

DRAZEN, J. M.; MORRISSEY, S.; MALINA, D.; HAMEL, M. B.; CAMPION, E. W. The Importance — and the Complexities — of Data Sharing. **New England Journal of Medicine**, v. 375, n. 12, p. 1182–1183, 22 set. 2016. DOI 10.1056/NEJMe1611027. Disponível em: <http://www.nejm.org/doi/10.1056/NEJMe1611027>. Acesso em: 9 set. 2020.

ESTADOS UNIDOS DAS AMÉRICAS. **Memorandum for the heads of executive departments and agencies: Open Data Policy: Managing Information as an Asset**. 2013. Disponível em: <https://obamawhitehouse.archives.gov/sites/default/files/omb/memoranda/2013/m-13-13.pdf>. Acesso em: 9 set. 2020.

EUROPEAN COMMISSION. **Guidelines on FAIR Data Management in Horizon 2020**. [S. l.]: European Commission, 26 jul. 2016. Disponível em: https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf. Acesso em: 9 set. 2020.

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, ago. 2010. DOI 10.1007/s10606-010-9117-8. Disponível em: <http://link.springer.com/10.1007/s10606-010-9117-8>. Acesso em: 9 set. 2020.

FECHER, B.; FRIESIKE, S. Open Science: One Term, Five Schools of Thought. *In*: BARTLING, S.; FRIESIKE, S. (orgs.). **Opening Science: the evolving guide on how the internet is changing research, collaboration and scholarly publishing**. Cham: Springer International Publishing, 2014. p. 17–47. DOI 10.1007/978-3-319-00026-8_2. Disponível em: http://link.springer.com/10.1007/978-3-319-00026-8_2. Acesso em: 9 set. 2020.

FERGUSON, L. How and why researchers share data (and why they don't). 3 nov. 2014. **The Wiley Network**. Disponível em: <https://www.wiley.com/network/researchers/licensing-and-open-access/how-and-why-researchers-share-data-and-why-they-dont>. Acesso em: 9 set. 2020.

FERREIRA, S. M. S. P.; TARGINO, M. das G. Apresentação. *In*: FERREIRA, S. M. S. P.; TARGINO, M. das G. L. (orgs.). **Acessibilidade e visibilidade de revistas científicas eletrônicas**. São Paulo, SP: Editora SENAC São Paulo : Cengage Learning, 2010. p. 15–18.

FISCHER, B. A.; ZIGMOND, M. J. The Essential Nature of Sharing in Science. **Science and Engineering Ethics**, v. 16, n. 4, p. 783–799, dez. 2010. DOI 10.1007/s11948-010-9239-x. Disponível em: <http://link.springer.com/10.1007/s11948-010-9239-x>. Acesso em: 9 set. 2020.

FOSTER. Facilitate Open Science Training for European Research. 2019. **FOSTER**. Disponível em: <https://www.fosteropenscience.eu/>. Acesso em: 9 set. 2020.

GARVEY, W. D. **Communication, the essence of science**: facilitating information exchange among librarians, scientists, engineers, and students. Oxford ; New York: Pergamon Press, 1979(Pergamon international library of science, technology, engineering, and social studies).

GARVEY, W. D.; GRIFFITH, B. C. Communication and information processing within scientific disciplines: Empirical findings for Psychology. **Information Storage and Retrieval**, v. 8, n. 3, p. 123–136, jun. 1972. DOI 10.1016/0020-0271(72)90041-1. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/0020027172900411>. Acesso em: 9 set. 2020.

GIGLIA, E.; SWAN, A. Open Access to data for a new, open science. **European Journal of Physical and Rehabilitation Medicine**, v. 48, n. 4, p. 713–716, dez. 2012. .

GILLILAND, A. J. Setting the stage. *In*: BACA, M. (org.). **Introduction to Metadata**. 3. ed. Los Angeles: Getty Publications, 2016. Disponível em: <http://www.getty.edu/publications/intrometadata/setting-the-stage/>. Acesso em: 6 set. 2020.

GÓMEZ, N.-D; MÉNDEZ, E.; HERNÁNDEZ-PÉREZ, T. Social sciences and humanities research data and metadata: A perspective from thematic data repositories. **El profesional de la información**, 2016, julio-agosto, v. 25, n. 4. eISSN: 1699-2407. Disponível em: <http://www.elprofesionaldelainformacion.com/contenidos/2016/jul/04.pdf>. Acesso em: 09 set. 2020.

GRAY, J. Jim Gray on e-Science: a transformed scientific method. *In*: HEY, T.; TANSLEY, S.; TOLLE, K. (orgs.). **The Fourth Paradigm: Data-Intensive Scientific Discovery**. [S. l.]: Microsoft Research, 2009. p. xvii–xxxii. Disponível em: <https://www.microsoft.com/en-us/research/publication/fourth-paradigm-data-intensive-scientific-discovery/>.

GUANAES, P. C. V. Quais direitos podem ser alegados sobre dados de pesquisa? *In*: SHINTAKU, M.; SALES, L. F.; COSTA, M. (orgs.). **Tópicos sobre dados abertos para editores científicos**. 1. ed. [S. l.]: ABEC, 2020. p. 79–85. DOI 10.21452/978-85-93910-04-3.cap7. Disponível em:

https://www.abecbrasil.org.br/arquivos/Topicos_dados_abertos_editores_cientificos.pdf#07. Acesso em: 9 set. 2020.

GUANAES, P. C. V.; GUIMARÃES, M. C. S. Modelos de gestão de revistas científicas: uma discussão necessária. **Perspectivas em Ciência da Informação**, v. 17, n. 1, p. 56–73, mar. 2012. DOI 10.1590/S1413-99362012000100004. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1413-99362012000100004&lng=pt&tlng=pt. Acesso em: 9 set. 2020.

GUIBAULT, L.; WIEBE, A. (Orgs.). **Safe to be open: study on the protection of research data and recommendations for access and usage**. Göttingen: Göttingen University Press, 2013. DOI 10.17875/gup2013-160. Disponível em: <http://resolver.sub.uni-goettingen.de/purl?univerlag-isbn-978-3-86395-147-4>. Acesso em: 9 set. 2020.

HARVEY, D. R. **Digital curation: a how-to-do-it manual**. New York: Neal-Schuman Publishers, 2010(How-to-do-it manuals, no. 170).

HATCHER, J. Implementing Open Data: The Open Data Commons Project. **Open Source Business Resource**, Ottawa, n. February 2008, 2008. .

HENDERSON, M. E. **Data management: a practical guide for librarians**. Lanham, Maryland: Rowman & Littlefield, 2017(Practical guides for librarians, no. 28).

HENNING, P. C.; RIBEIRO, C. J. S.; SALES, L. F.; MOREIRA, L. R.; SANTOS, L. O. B. da S. Desmistificando os princípios FAIR: conceitos, métricas, tecnologias e aplicações inseridas no ecossistema dos dados FAIR. **Pesquisa Brasileira em Ciência da Informação e Biblioteconomia**, 25 set. 2019. DOI 10.22478/ufpb.1981-0695.2019v14n3.46969. Disponível em: <https://periodicos.ufpb.br/index.php/abcib/article/view/46969>. Acesso em: 9 set. 2020.

HENNING, P. C.; RIBEIRO, C. J. S.; SANTOS, L. O. B. D. S.; SANTOS, P. X. D. GO FAIR e os princípios FAIR: o que representam para a expansão dos dados de pesquisa no âmbito da Ciência Aberta. **Em Questão**, v. 25, n. 2, p. 389–412, 26 abr. 2019. DOI 10.19132/1808-5245252.389-412. Disponível em: <https://seer.ufrgs.br/EmQuestao/article/view/84753>. Acesso em: 9 set. 2020.

HEY, T.; HEY, J. e-Science and its implications for the library community. **Library Hi Tech**, v. 24, n. 4, p. 515–528, out. 2006. DOI 10.1108/07378830610715383. Disponível em: <https://www.emerald.com/insight/content/doi/10.1108/07378830610715383/full/html>. Acesso em: 9 set. 2020.

HEY, T.; TREFETHEN, A. E. The UK e-Science Core Programme and the Grid. **Future Generation Computer Systems**, v. 18, n. 8, p. 1017–1031, out. 2002. DOI 10.1016/S0167-739X(02)00082-1. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0167739X02000821>. Acesso em: 9 set. 2020.

HOUAISS, A. **Dicionário Houaiss da língua portuguesa**. Rio de Janeiro: [s. n.], 2019. Disponível em: https://houaiss.uol.com.br/corporativo/apps/uol_www/v5-4/html/index.php#0. Acesso em: 9 set. 2020.

HURD, J. M. The transformation of scientific communication: A model for 2020. **Journal of the American Society for Information Science**, v. 51, n. 14, p. 1279–1283, 2000. DOI

10.1002/1097-4571(2000)9999:9999<:AID-ASI1044>3.0.CO;2-1. Disponível em: <https://asistdl.onlinelibrary.wiley.com/doi/abs/10.1002/1097-4571%282000%299999%3A9999%3C%3A%3AAID-ASI1044%3E3.0.CO%3B2-1>. Acesso em: 9 set. 2020.

IMMING, M. **Fair data advanced use cases: from principles to practice in the netherlands**. [S. l.]: Zenodo, 23 abr. 2018. DOI 10.5281/ZENODO.1250535. Disponível em: <https://zenodo.org/record/1250535>. Acesso em: 9 set. 2020.

INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA. Repositórios Digitais. 2019. Disponível em: <https://ibict.br/informacao-para-a-pesquisa/repositorios-digitais>. Acesso em: 9 set. 2020.

INTERNATIONAL DOI FOUNDATION. **DOI® Handbook**. [S. l.]: INTERNATIONAL DOI FOUNDATION, 2019. Disponível em: <https://www.doi.org/hb.html>. Acesso em: 9 set. 2020.

KAUARK, F. da S.; MANHÃES, F. C.; MEDEIROS, C. H. **Metodologia da pesquisa : guia prático**. Itabuna: Via Litterarum, 2010. Disponível em: http://www.pgcl.uenf.br/arquivos/livrodemetodologiadapesquisa2010_011120181549.pdf. Acesso em: 9 set. 2020.

KELLAM, L. M.; THOMPSON, K. (Orgs.). **Databrarianship: the academic data librarian in theory and practice**. Chicago, Illinois: Association of College and Research Libraries, a division of the American Library Association, 2016.

KIM, Y.; STANTON, J. M. Institutional and individual factors affecting scientists' data-sharing behaviors: A multilevel analysis. **Journal of the Association for Information Science and Technology**, v. 67, n. 4, p. 776–799, abr. 2016. DOI 10.1002/asi.23424. Disponível em: <http://doi.wiley.com/10.1002/asi.23424>. Acesso em: 9 set. 2020.

KOLTAY, T. Data literacy for researchers and data librarians. **Journal of Librarianship and Information Science**, v. 49, n. 1, p. 3–14, mar. 2017. DOI 10.1177/0961000615616450. Disponível em: <http://journals.sagepub.com/doi/10.1177/0961000615616450>. Acesso em: 9 set. 2020.

KUHN, T. S. Reflections on my Critics. *In*: LAKATOS, I.; MUSGRAVE, A. (orgs.). **Criticism and the Growth of Knowledge**. 1. ed. [S. l.]: Cambridge University Press, 1970. p. 231–278. DOI 10.1017/CBO9781139171434.011. Disponível em: https://www.cambridge.org/core/product/identifier/CBO9781139171434A016/type/book_p art. Acesso em: 9 set. 2020.

KUM, H.-C.; AHALT, S.; CARSEY, T. M. Dealing with Data: Governments Records. **Science**, v. 332, n. 6035, p. 1263–1263, 10 jun. 2011. DOI 10.1126/science.332.6035.1263-a. Disponível em: <https://www.sciencemag.org/lookup/doi/10.1126/science.332.6035.1263-a>. Acesso em: 9 set. 2020.

LABASTIDA, I.; MARGONI, T. Licensing FAIR Data for Reuse. **Data Intelligence**, v. 2, n. 1–2, p. 199–207, jan. 2020. DOI 10.1162/dint_a_00042. Disponível em: https://www.mitpressjournals.org/doi/abs/10.1162/dint_a_00042. Acesso em: 18 set. 2020.

LARIVIÈRE, V.; HAUSTEIN, S.; MONGEON, P. The Oligopoly of Academic Publishers in the Digital Era. **PLOS ONE**, v. 10, n. 6, p. e0127502, 10 jun. 2015. DOI 10.1371/journal.pone.0127502. Disponível em: <https://dx.plos.org/10.1371/journal.pone.0127502>. Acesso em: 9 set. 2020.

LECLERE, F. Too Many Researchers Are Reluctant to Share Their Data. **CHE**, , seq. The Review, 3 ago. 2010. Disponível em: <https://www.chronicle.com/article/too-many-researchers-are-reluctant-to-share-their-data/>. Acesso em: 9 set. 2020.

LEE, C.; TIBBO, H. Digital Curation and Trusted Repositories: Steps Toward Success. **Journal of Digital Information**, v. 8, n. 2, 2007. Disponível em: <https://journals.tdl.org/jodi/index.php/jodi/article/view/229>. Acesso em: 9 set. 2020.

LEITE, F. C. L. **Como gerenciar e ampliar a visibilidade da informação científica brasileira repositórios institucionais de acesso aberto**. Brasília, DF: Ibict, 2009. Disponível em: <http://livroaberto.ibict.br/handle/1/775>. Acesso em: 9 set. 2020.

LEITE, Fernando César Lima. **Gestão do conhecimento científico no contexto acadêmico: proposta de um modelo conceitual**. 2006. 240 f. Dissertação (Mestrado) – Universidade de Brasília, Faculdade de Economia, Administração, Contabilidade e Ciência da Informação e Documentação, Departamento de Ciência da Informação e Documentação, Brasília, 2006. Disponível em: <https://repositorio.unb.br/handle/10482/3975>. Acesso em: 9 set. 2020.

LEITE, Fernando César Lima. **Modelo genérico de gestão da informação científica para instituições de pesquisa na perspectiva da comunicação científica e do acesso aberto**. 2011. 262 f. Tese (Doutorado em Ciência da Informação) – Universidade de Brasília, Brasília, 2011. Disponível em: <https://repositorio.unb.br/handle/10482/3975>. Acesso em: 9 set. 2020.

LEMONS, A. **Cibercultura: tecnologia e vida social na cultura contemporânea**. Porto Alegre: Sulina, 2002.

LÉVY, P. **Cibercultura**. trad. Carlos Irineu Da Costa. 3. ed. São Paulo: Editora34, 2010.

LONGO, D. L.; DRAZEN, J. M. Data Sharing. **New England Journal of Medicine**, v. 374, n. 3, p. 276–277, 21 jan. 2016. DOI 10.1056/NEJMe1516564. Disponível em: <http://www.nejm.org/doi/10.1056/NEJMe1516564>. Acesso em: 9 set. 2020.

LYNCH, C. A. Institutional Repositories: Essential Infrastructure For Scholarship In The Digital Age. **Libraries and the Academy**, v. 3, n. 2, p. 327–336, 2003. DOI 10.1353/pla.2003.0039. Disponível em: http://muse.jhu.edu/content/crossref/journals/portal_libraries_and_the_academy/v003/3.2lynch.html. Acesso em: 9 set. 2020.

LYON, L. **Dealing with Data: roles, rights, responsibilities and relationships**. consultancy report. Inglaterra: UKOLN, jun. 2007. Disponível em: <http://www.ukoln.ac.uk/ukoln/staff/e.j.lyon/publications.html#2007-06-19>. Acesso em: 9 set. 2020.

- MACHADO, J. Dados abertos e ciência aberta. *In*: ALBAGLI, S.; MACIEL, M. L.; ABDO, A. H. (orgs.). **Ciência aberta, questões abertas**. Brasília: IBICT, 2015. p. 201–228. Disponível em: <http://livroaberto.ibict.br/handle/1/1060>. Acesso em: 9 set. 2020.
- MALAPELA, T. **Guidelines on FAIR Data Management in Horizon 2020**. 24 ago. 2016. **AIMS**. Disponível em: <http://aims.fao.org/activity/blog/guidelines-fair-data-management-horizon-2020>. Acesso em: 9 set. 2020.
- MARCHIORI, P. Z.; ADAMI, A. Autoria e leitura de artigos por docentes pesquisadores: motivações e barreiras. *In*: FERREIRA, S. M. S. P.; TARGINO, M. das G. (orgs.). **Preparação de revistas científicas: teoria e prática**. São Paulo: Reichmann & Autores Ed, 2005. p. 73–100.
- MARQUES, F. Ciência transparente. n. 218, 2014. Disponível em: <https://revistapesquisa.fapesp.br/ciencia-transparente/>. Acesso em: 9 set. 2020.
- MARTINS, R. B. **Do papel ao digital**: a trajetória de duas revistas científicas brasileiras. 2003. 182 f. Dissertação (Mestrado em Ciência da Informação) – Universidade Federal do Rio de Janeiro / Instituto Brasileiro de Informação em Ciência e Tecnologia, Rio de Janeiro, 2003. Disponível em: <http://ridi.ibict.br/handle/123456789/716>. Acesso em: 9 set. 2020.
- MAYERNIK, M. S. **Metadata realities for cyberinfrastructure: data authors as metadata creators**. 2011. 337 f. Dissertation (Doctor of Philosophy in Information Studies) – University of California, Los Angeles, 2011. DOI: 10.2139/ssrn.2042653. Disponível em: <https://ssrn.com/abstract=2042653>. Acesso em: 9 set. 2020.
- MEADOWS, A. To Share or not to Share? That is the (Research Data) Question.... **The Scholarly Kitchen**, 11 nov. 2014. Disponível em: <https://scholarlykitchen.sspnet.org/2014/11/11/to-share-or-not-to-share-that-is-the-research-data-question/>. Acesso em: 9 set. 2020.
- MEADOWS, A. J. **A comunicação científica**. Brasília: Briquet de Lemos/livros, 1999.
- MEDEIROS, J. D. S.; CAREGNATO, S. E. Compartilhamento de dados e e-Science: explorando um novo conceito para a comunicação científica. **Liinc em Revista**, v. 8, n. 2, 18 dez. 2012. DOI 10.18617/liinc.v8i2.488. Disponível em: <http://revista.ibict.br/liinc/article/view/3368>. Acesso em: 9 set. 2020.
- MENNES, M.; BISWAL, B. B.; CASTELLANOS, F. X.; MILHAM, M. P. Making data sharing work: The FCP/INDI experience. **NeuroImage**, v. 82, p. 683–691, nov. 2013. DOI 10.1016/j.neuroimage.2012.10.064. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S1053811912010671>. Acesso em: 9 set. 2020.
- MOLLOY, J. C. The Open Knowledge Foundation: Open Data Means Better Science. **PLoS Biology**, v. 9, n. 12, p. e1001195, 6 dez. 2011. DOI 10.1371/journal.pbio.1001195. Disponível em: <https://dx.plos.org/10.1371/journal.pbio.1001195>. Acesso em: 9 set. 2020.
- MONS, B. **Data stewardship for open science: implementing FAIR principles**. 1. ed. [S. l.]: Chapman and Hall/CRC, 2018. DOI 10.1201/9781315380711. Disponível em: <https://www.taylorfrancis.com/books/9781498753180>. Acesso em: 9 set. 2020.

MORENO, F. P. Repositórios de dados de pesquisa na Espanha: breve análise. **Encontros Bibli**: revista eletrônica de biblioteconomia e ciência da informação, v. 23, n. 53, p. 52–63, 6 set. 2018. DOI 10.5007/1518-2924.2018v23n53p52. Disponível em: <https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2018v23n53p52>. Acesso em: 9 set. 2020.

MORRISON, H. G. **Freedom for scholarship in the internet age**. 2012. (Dissertation) Ph.D. – Communication, Art & Technology: School of Communication, 2012. Disponível em: <https://summit.sfu.ca/item/12537>. Acesso em: 9 set. 2020.

MUELLER, S. P. M.; PASSOS, E. J. L. As questões da comunicação científica e a ciência da informação. *In*: MUELLER, S.P.M; PASSOS, E. J. L. (orgs.). **Comunicação científica estudos avançados em ciência da informação**. Brasília: Universidade de Brasília, 2000. p. 13–22.

MUELLER, Suzana Pinheiro Machado. O periódico científico. *In*: CAMPELLO, B. S.; CENDÓN, B. V.; KREMER, J. M. (orgs.). **Fontes de informação para pesquisadores e profissionais**. Coleção Aprender. Belo Horizonte: Editora UFMG, 2000. p. 73–96.

MUELLER-LANGER, F.; ANDREOLI-VERSBACH, P. Open access to research data: Strategic delay and the ambiguous welfare effects of mandatory data disclosure. **Information Economics and Policy**, v. 42, p. 20–34, mar. 2018. DOI 10.1016/j.infoecopol.2017.05.004. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0167624516301147>. Acesso em: 9 set. 2020.

NASSI-CALÒ, L. Projeto Making Data Count incentiva compartilhamento de dados de pesquisa [online]. **SciELO em Perspectiva**, 2015. Disponível em: <https://blog.scielo.org/blog/2015/10/01/projeto-making-data-count-incentiva-compartilhamento-de-dados-de-pesquisa/>. Acesso em: 9 set. 2020.

NATIONAL ACADEMIE OF SCIENCES. **A question of balance**: private rights and the public interest in scientific and technical databases. Washington, D.C.: National Academies Press, 1999. p. 9692(9692). DOI 10.17226/9692. Disponível em: <http://www.nap.edu/catalog/9692>. Acesso em: 9 set. 2020.

NATIONAL INSTITUTE OF HEALTH. Principles and Guidelines for Reporting Preclinical Research. 13 ago. 2015. **National Institutes of Health (NIH)**. Disponível em: <https://www.nih.gov/research-training/rigor-reproducibility/principles-guidelines-reporting-preclinical-research>. Acesso em: 9 set. 2020.

NATIONAL SCIENCE FOUNDATION. **Cyberinfrastructure Vision for 21st Century Discovery**. EUA: NSF, 2007(NSF 07-28, 07–28). Disponível em: https://www.nsf.gov/publications/pub_summ.jsp?ods_key=nsf17001&org=NSF. Acesso em: 9 set. 2020.

NATURE. **Scientific Data**: principles. 2019. Disponível em: <https://www.nature.com/sdata/about/principles>. Acesso em: 9 set. 2020.

NEUMANN, J.; BRASE, J. DataCite and DOI names for research data. **Journal of Computer-Aided Molecular Design**, v. 28, n. 10, p. 1035–1041, out. 2014. DOI 10.1007/s10822-014-9776-5. Disponível em: <http://link.springer.com/10.1007/s10822-014-9776-5>. Acesso em: 9 set. 2020.

NEUROTH, H.; STRATHMANN, S.; OSSWALD, A.; LUDWIG, J. (Orgs.). **Digital curation of research data: experiences of a baseline study in Germany**. Göttingen: VWH, 2013.

NIELSEN, M. A. **Reinventing discovery: the new era of networked science**. [S. l.: s. n.], 2014. Disponível em:
<http://www.vlebooks.com/vleweb/product/openreader?id=none&isbn=9781400839452>.
 Acesso em: 9 set. 2020.

NIU, J. **Perceived Documentation Quality of Social Science Data**. 2009. Dissertation – University of Michigan, Michigan, 2009. Disponível em:
<http://hdl.handle.net/2027.42/63871>. Acesso em: 9 set. 2020.

OECD. **OECD Principles and Guidelines for Access to Research Data from Public Funding**. [S. l.]: OECD Publishing, 2007. DOI 10.1787/9789264034020-en-fr. Disponível em: https://www.oecd-ilibrary.org/science-and-technology/oecd-principles-and-guidelines-for-access-to-research-data-from-public-funding_9789264034020-en-fr. Acesso em: 11 set. 2020.

OPEN ACCESS INFRASTRUCTURE FOR RESEARCH IN EUROPE (OPENAIRE). What is metadata for research data? 27 set. 2018. **OpenAIRE**. Disponível em:
<https://www.openaire.eu/what-is-metadata>. Acesso em: 24 set. 2020.

OPEN KNOWLEDGE FOUNDATION. **Definição de Conhecimento Aberto**. 2019. Disponível em: <https://opendefinition.org/od/2.0/pt-br/>. Acesso em: 9 set. 2020.

PADILLA NAVARRO, P. A.; CURAQUEO PICHIHUECHE, O. J.; CANCINO SALAS, R. D.; GATICA ÁLVAREZ, M. H. Acceso a datos de investigación e información científica en Chile. **Revista española de Documentación Científica**, v. 36, n. 3, p. en010, 30 set. 2013. DOI 10.3989/redc.2013.3.960. Disponível em:
<http://redc.revistas.csic.es/index.php/redc/article/view/812/983>. Acesso em: 11 set. 2020.

PAMPEL, H.; VIERKANT, P.; SCHOLZE, F.; BERTELMANN, R.; KINDLING, M.; KLUMP, J.; GOEBELBECKER, H.-J.; GUNDLACH, J.; SCHIRMBACHER, P.; DIEROLF, U. Making Research Data Repositories Visible: The re3data.org Registry. **PLoS ONE**, v. 8, n. 11, p. e78080, 4 nov. 2013. DOI 10.1371/journal.pone.0078080. Disponível em: <https://dx.plos.org/10.1371/journal.pone.0078080>. Acesso em: 9 set. 2020.

PARHAM, S. W.; BODNAR, J.; FUCHS, S. Supporting tomorrow's research: Assessing faculty data curation needs at Georgia Tech. **College & Research Libraries News**, v. 73, n. 1, p. 10–13, 1 jan. 2012. DOI 10.5860/crln.73.1.8686. Disponível em:
<http://crln.acrl.org/index.php/crlnews/article/view/8686>. Acesso em: 11 set. 2020.

PENEV, L.; CHAVAN, V.; GEORGIEV, T.; STOEV, P. **Data papers as incentives for opening biodiversity data: one year of experience and perspectives for the future**. 2013. Disponível em:
http://eubon.eu/news/4173_Data%20Papers%20as%20Incentives%20for%20Opening%20Biodiversity%20Data:%20One%20Year%20of%20Experience%20and%20Perspectives%20for%20The%20Future. Acesso em: 19 set. 2020.

PINHEIRO, L. V. R. Do acesso livre à ciência aberta: conceitos e implicações na comunicação científica. **RECIIS**, v. 8, n. 2, p. 153–165, 2014. Disponível em:

<https://www.reciis.icict.fiocruz.br/index.php/reciis/article/viewArticle/629/1090>. Acesso em: 9 set. 2020.

PINTO, C. S.; COSTA, J. L. Padrões de comunicação em diferentes comunidades científicas. *In*: COSTA, S. M. de S.; LEITE, F. C. L.; TAVARES, R. B. (orgs.). **Comunicação da informação, gestão da informação e gestão do conhecimento**. Brasília, DF: IBICT, 2018. p. 145–159. Disponível em: <http://livroaberto.ibict.br/handle/123456789/1071>. Acesso em: 9 set. 2020.

PIORUN, M. **E-Science as a catalyst for transformational change in university research libraries**: a dissertation. 2013. Thesis (Ph.D.) – University of Massachusetts Medical School, 2013. DOI 10.13028/44A8-4171. Disponível em: https://escholarship.umassmed.edu/lib_articles/146/. Acesso em: 9 set. 2020.

PIWOWAR, H. A. Who Shares? Who Doesn't? Factors Associated with Openly Archiving Raw Research Data. **PLoS ONE**, v. 6, n. 7, p. e18657, 13 jul. 2011. DOI 10.1371/journal.pone.0018657. Disponível em: <https://dx.plos.org/10.1371/journal.pone.0018657>. Acesso em: 9 set. 2020.

PIWOWAR, H. A.; CHAPMAN, W. W. Identifying data sharing in biomedical literature. **AMIA: Annual Symposium proceedings**, p. 596–600, 6 nov. 2008. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/18998887/>. Acesso em: 9 set. 2020.

PIWOWAR, H. A.; DAY, R. S.; FRIDSMA, D. B. Sharing Detailed Research Data Is Associated with Increased Citation Rate. **PLoS ONE**, v. 2, n. 3, p. e308, 21 mar. 2007. DOI 10.1371/journal.pone.0000308. Disponível em: <https://dx.plos.org/10.1371/journal.pone.0000308>. Acesso em: 9 set. 2020.

PIWOWAR, H. A.; VISION, T. J. Data reuse and the open data citation advantage. **PeerJ**, v. 1, p. e175, 1 out. 2013. DOI 10.7717/peerj.175. Disponível em: <https://peerj.com/articles/175>. Acesso em: 9 set. 2020.

PONTIKA, N.; KNOTH, P.; CANCELLIERI, M.; PEARCE, S. Fostering open science to research using a taxonomy and an eLearning portal. *In*: THE 15TH INTERNATIONAL CONFERENCE, 2015. **Proceedings** [...]. Graz, Austria: ACM Press, 2015. p. 1–8. DOI 10.1145/2809563.2809571. Disponível em: <http://dl.acm.org/citation.cfm?doid=2809563.2809571>. Acesso em: 9 set. 2020.

RAUTENBERG, S.; SOUZA, L. de. Um estudo para representação de dados abertos de editoração científica na web de dados. *In*: FARIAS SALES, L.; COSTA, M. **Tópicos sobre dados abertos para editores científicos**. 1. ed. Botucatu: ABEC, 2020. p. 47–63. DOI 10.21452/978-85-93910-04-3.cap5. Disponível em: https://www.abecbrasil.org.br/arquivos/Topicos_dados_abertos_editores_cientificos.pdf#05. Acesso em: 9 set. 2020.

RE3DATA. Re3data: Browse by country. 2020. Disponível em: <https://www.re3data.org/browse/by-country/>. Acesso em: 15 set. 2020.

RIBES, D.; LEE, C. P. Sociotechnical Studies of Cyberinfrastructure and e-Research: Current Themes and Future Trajectories. **Computer Supported Cooperative Work (CSCW)**, v. 19, n. 3–4, p. 231–244, ago. 2010. DOI 10.1007/s10606-010-9120-0.

Disponível em: <http://link.springer.com/10.1007/s10606-010-9120-0>. Acesso em: 11 set. 2020.

ROBINSON-GARCIA, N.; MONGEON, P.; JENG, W.; COSTAS, R. DataCite as a novel bibliometric source: Coverage, strengths and limitations. **Journal of Informetrics**, v. 11, n. 3, p. 841–854, ago. 2017. DOI 10.1016/j.joi.2017.07.003. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S1751157717300834>. Acesso em: 25 set. 2020.

RODRIGUES, Eloy; SARAIVA, R.; RIBEIRO, C.; FERNANDES, E. M. **Os repositórios de dados científicos: estado da arte**. Minho, Portugal: Universidade do Minho, 2010. Disponível em: <https://repositorio-aberto.up.pt/handle/10216/23806>. Acesso em: 9 set. 2020.

ROSENDAAL, H.; GEURTS, P. Forces and functions in scientific communication : an analysis of their interplay. *In*: COOPERATIVE RESEARCH INFORMATION SYSTEMS IN PHYSICS, 1997. **Anais [...]**. [S. l.: s. n.], 1997. Disponível em: <http://www.physik.uni-oldenburg.de/conferences/crisp97/roosendaal.html>. Acesso em: 9 set. 2020.

RUEDA, L.; FENNER, M.; CRUSE, P. DataCite: Lessons Learned on Persistent Identifiers for Research Data. **International Journal of Digital Curation**, v. 11, n. 2, p. 39–47, 4 jul. 2017. DOI 10.2218/ijdc.v11i2.421. Disponível em: <http://www.ijdc.net/article/view/11.2.39>. Acesso em: 25 set. 2020.

RUTTENBERG, S.; RISHBETH, H. World Data Centres: past, present and future. **Journal of Atmospheric and Terrestrial Physics**, v. 56, n. 7, p. 865–870, maio 1994. DOI 10.1016/0021-9169(94)90148-1. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/0021916994901481>. Acesso em: 9 set. 2020.

SALES, L. F. **Integração semântica de publicações científicas e dados de pesquisa: proposta de modelo de publicação ampliada para a área de Ciências Nucleares**. 2014. Tese (Doutorado em Ciência da Informação) – Instituto Brasileiro de Informação em Ciência e Tecnologia/Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2014. Disponível em: <http://ridi.ibict.br/handle/123456789/874>. Acesso em: 9 set. 2020.

SALES, Luana Farias; SAYÃO, L. F. Ampliando as fronteiras da editoração científica: o papel dos repositórios de dados. *In*: SHINTAKU, M.; SALES, L. F.; COSTA, M. (orgs.). **Tópicos sobre dados abertos para editores científicos**. 1. ed. [S. l.]: ABEC, 2020. p. 39–45. DOI 10.21452/978-85-93910-04-3.cap4. Disponível em: https://www.abecbrasil.org.br/arquivos/Topicos_dados_abertos_editores_cientificos.pdf#04. Acesso em: 9 set. 2020.

SALES, Luana Farias; SAYÃO, L. F. O impacto da curadoria digital dos dados de pesquisa na Comunicação Científica. **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação**, v. 0, n. 0, p. 118–135, 18 dez. 2012. DOI 10.5007/1518-2924.2012v17nesp2p118. Disponível em: <http://www.periodicos.ufsc.br/index.php/eb/article/view/27417>. Acesso em: 9 set. 2020.

SANTAREM SEGUNDO, J. E. Web semântica e linked data como aliados na consolidação da Ciência Aberta. *In*: FARIAS SALES, L.; COSTA, M. **Tópicos sobre dados abertos para editores científicos**. 1. ed. [S. l.]: ABEC, 2020. p. 65–78. DOI 10.21452/978-85-93910-04-3.cap6. Disponível em:

https://www.abecbrasil.org.br/arquivos/Topicos_dados_abertos_editores_cientificos.pdf#06. Acesso em: 9 set. 2020.

SAVAGE, C. J.; VICKERS, A. J. Empirical Study of Data Sharing by Authors Publishing in PLoS Journals. **PLoS ONE**, v. 4, n. 9, p. e7078, 18 set. 2009. DOI 10.1371/journal.pone.0007078. Disponível em: <https://dx.plos.org/10.1371/journal.pone.0007078>. Acesso em: 9 set. 2020.

SAYÃO, L. F.; SALES, L. F. Algumas considerações sobre os repositórios digitais de dados de pesquisa. **Informação & Informação**, v. 21, n. 2, p. 90, 2016. DOI 10.5433/1981-8920.2016v21n2p90. Disponível em: <http://www.uel.br/revistas/uel/index.php/informacao/article/view/27939>. Acesso em: 9 set. 2020.

SAYÃO, L. F.; SALES, L. F. Curadoria digital e dados de pesquisa. **AtoZ: novas práticas em informação e conhecimento**, v. 5, n. 2, p. 67, 2016. DOI 10.5380/atoz.v5i2.49708. Disponível em: <http://revistas.ufpr.br/atoz/article/view/49708>. Acesso em: 9 set. 2020.

SAYÃO, L. F.; SALES, L. F. Curadoria digital: um novo patamar para preservação de dados digitais de pesquisa. **Informação & Sociedade: Estudos**, v. 22, n. 3, out. 2012. Disponível em: <https://periodicos.ufpb.br/ojs/index.php/ies/article/view/12224>. Acesso em: 9 set. 2020.

SAYÃO, L. F.; SALES, L. F. **Guia de gestão de dados de pesquisa para bibliotecários e pesquisadores**. Rio de Janeiro: CNEN/IEN, 2015. Disponível em: <http://www.cnen.gov.br/component/content/article/75-cin/material-didatico-cnen/160-guia-de-gestao-de-dados-de-pesquisa>. Acesso em: 9 set. 2020.

SCIENTIFIC ELECTRONIC LIBRARY ONLINE. Princípios reitores FAIR publicados em periódico do Nature Publishing Group [online]. **SciELO em Perspectiva**, 2016. Disponível em: <https://blog.scielo.org/blog/2016/03/16/principios-orientadores-fair-publicados-em-periodico-do-nature-publishing-group/>. Acesso em: 9 set. 2020.

SEMELER, A. R.; PINTO, A. L. Os diferentes conceitos de dados de pesquisa na abordagem da biblioteconomia de dados. **Ciência da Informação**, v. 48, n. 1, maio 2019. Disponível em: <http://revista.ibict.br/ciinf/article/view/4461>. Acesso em: 9 set. 2020.

SHINTAKU, M.; COSTA, S. M. de S. A comunicação científica entre pesquisadores da surdez do ponto de vista da linguagem. In: COSTA, S. M. de S.; LEITE, F. C. L.; TAVARES, R. B. (orgs.). **Comunicação da informação, gestão da informação e gestão do conhecimento**. Brasília, DF: IBICT, 2018. p. 161–178. Disponível em: <http://livroaberto.ibict.br/handle/123456789/1071>. Acesso em: 9 set. 2020.

SHINTAKU, M.; FERREIRA JUNIOR, R. S. Gestão de dados em periódicos científicos. In: SALES, Luana Farias; COSTA, M. **Tópicos sobre dados abertos para editores científicos**. 1. ed. [S. l.]: ABEC, 2020. p. 231–240. DOI 10.21452/978-85-93910-04-3.cap18. Disponível em: https://www.abecbrasil.org.br/arquivos/Topicos_dados_abertos_editores_cientificos.pdf#18. Acesso em: 9 set. 2020.

SOEHNER, C.; STEEVES, C.; WARD, J. **E-Science and data support services: a study of ARL Member Institutions**. Washington, DC: Association of Research Libraries, 2010. Disponível em: <https://eric.ed.gov/?id=ED528643>. Acesso em: 9 set. 2020.

SØNDERGAARD, T. F.; ANDERSEN, J.; HJØRLAND, B. 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, jun. 2003. DOI 10.1108/00220410310472509. Disponível em: <https://www.emerald.com/insight/content/doi/10.1108/00220410310472509/full/html>. Acesso em: 9 set. 2020.

SONNENWALD, D. H. Scientific collaboration. **Annual Review of Information Science and Technology**, v. 41, n. 1, p. 643–681, 2007. DOI 10.1002/aris.2007.1440410121. Disponível em: <http://doi.wiley.com/10.1002/aris.2007.1440410121>. Acesso em: 9 set. 2020.

SOUZA, R. F. de. Universo de ciência e tecnologia: organização e representação em classificações do conhecimento. *In*: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 13., 2012. **Anais [...]**. Rio de Janeiro: ENANCIB, 2012. v. 13, . Disponível em: <http://hdl.handle.net/123456789/641>. Acesso em: 9 set. 2020.

SPECKA, X.; GÄRTNER, P.; HOFFMANN, C.; SVOBODA, N.; STECKER, M.; EINSPANIER, U.; SENKLER, K.; ZOARDER, M. A. M.; HEINRICH, U. The BonaRes metadata schema for geospatial soil-agricultural research data – Merging INSPIRE and DataCite metadata schemes. **Computers & Geosciences**, v. 132, p. 33–41, nov. 2019. DOI 10.1016/j.cageo.2019.07.005. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S009830041930086X>. Acesso em: 25 set. 2020.

STARR, J.; GASTL, A. isCitedBy: A Metadata Scheme for DataCite. **D-Lib Magazine**, v. 17, n. 1/2, jan. 2011. DOI 10.1045/january2011-starr. Disponível em: <http://www.dlib.org/dlib/january11/starr/01starr.html>. Acesso em: 9 set. 2020.

STEINHART, Gail. Libraries as Distributors of Geospatial Data: Data Management Policies as Tools for Managing Partnerships. **Library Trends**, v. 55, n. 2, p. 264–284, 2006. DOI 10.1353/lib.2006.0063. Disponível em: http://muse.jhu.edu/content/crossref/journals/library_trends/v055/55.2steinhart.html. Acesso em: 9 set. 2020.

STODDEN, V.; GUO, P.; MA, Z. Toward Reproducible Computational Research: An Empirical Analysis of Data and Code Policy Adoption by Journals. **PLoS ONE**, v. 8, n. 6, p. e67111, 21 jun. 2013. DOI 10.1371/journal.pone.0067111. Disponível em: <https://dx.plos.org/10.1371/journal.pone.0067111>. Acesso em: 11 set. 2020.

STRAUSS, A. L.; CORBIN, J. **Pesquisa qualitativa técnicas e procedimentos para o desenvolvimento de teoria fundamentada**. Porto Alegre: Artmed, 2008.

STUART, D.; BAYNES, G.; HRYNASZKIEWICZ, I.; ALLIN, K.; PENNY, D.; MITHU LUCRAFT; ASTELL, M. Whitepaper: Practical challenges for researchers in data sharing. , p. 698462 Bytes, 2018. DOI 10.6084/M9.FIGSHARE.5975011. Disponível em: https://figshare.com/articles/Whitepaper_Practical_challenges_for_researchers_in_data_sharing/5975011. Acesso em: 9 set. 2020.

STUMPF, I. R. C. Passado e futuro das revistas científicas. **Ciência da Informação**, v. 25, n. 3, 1996. Disponível em: <http://revista.ibict.br/ciinf/article/view/637>. Acesso em: 9 set. 2020.

SWAN, A. **Diretrizes para políticas de desenvolvimento e promoção do acesso aberto**. trad. Patrícia Zimbres; Romes de Sousa Ramos. Brasília: UNESCO;IBICT, 2016(Série diretrizes para o acesso aberto). Disponível em: <https://unesdoc.unesco.org/ark:/48223/pf0000246018?posInSet=2&queryId=127e6c6c-fce7-4c7f-a961-71ee84b004c9>. Acesso em: 20 dez. 2019.

SWAN, A.; BROWN, S. **To share or not to share: publication and quality assurance of research data outputs: a report commissioned by the Research Information Network**. Project Report. [S. l.]: s.n., jun. 2008. Disponível em: <https://eprints.soton.ac.uk/266742/>. Acesso em: 9 set. 2020.

TENOPIR, C.; ALLARD, S.; DOUGLASS, K.; AYDINOGLU, A. U.; WU, L.; READ, E.; MANOFF, M.; FRAME, M. Data Sharing by Scientists: Practices and Perceptions. **PLoS ONE**, v. 6, n. 6, p. e21101, 29 jun. 2011. DOI 10.1371/journal.pone.0021101. Disponível em: <https://dx.plos.org/10.1371/journal.pone.0021101>. Acesso em: 11 set. 2020.

TENOPIR, C.; KING, D. W. A importância dos periódicos para o trabalho científico. **Revista de Biblioteconomia de Brasília**, v. 25, n. 1, p. 15–26, 2001. Disponível em: <https://brapci.inf.br/index.php/res/v/70824>. Acesso em: 9 set. 2020.

TORRES-SALINAS, D.; ROBINSON-GARCIA, N.; CABEZAS-CLAVIJO, Á. Compartir los datos de investigación en ciencia: introducción al data sharing. **El Profesional de la Información**, v. 21, n. 2, 2012. Disponível em: <http://eprints.rclis.org/16786/>. Acesso em: 9 set. 2020.

UNESCO. **Carta sobre la preservación del patrimonio digital**: UNESCO. 2003. Disponível em: http://portal.unesco.org/es/ev.php-URL_ID=17721&URL_DO=DO_TOPIC&URL_SECTION=201.html. Acesso em: 9 set. 2020.

UNESCO. **International standard classification of education**: ISCED 2011. Montreal, Quebec: UNESCO Institute for Statistics, 2012. Disponível em: <http://uis.unesco.org/sites/default/files/documents/international-standard-classification-of-education-isced-2011-en.pdf>. Acesso em: 9 set. 2020.

UNESCO. **UNISIST**: study report on the feasibility of a World Science Information System. Paris: UNESCO, 1971. Disponível em: <https://unesdoc.unesco.org/ark:/48223/pf0000064862>.

VASILEVSKY, N. A.; MINNIER, J.; HAENDEL, M. A.; CHAMPIEUX, R. E. Reproducible and reusable research: are journal data sharing policies meeting the mark? **PeerJ**, v. 5, p. e3208, 25 abr. 2017. DOI 10.7717/peerj.3208. Disponível em: <https://peerj.com/articles/3208>. Acesso em: 9 set. 2020.

VEIGA, V. **Ciência aberta e gestão de dados de pesquisa**. Rio de Janeiro: [s. n.], 2019. Disponível em: <https://www.arca.fiocruz.br/handle/icict/36361>. Acesso em: 9 set. 2020.

WALLIMAN, N. **Your research project: a step-by-step guide for the first-time researcher**. 2nd ed. London ; Thousand Oaks, Calif: Sage Publications, 2005.

WALPORT, M.; BREST, P. Sharing research data to improve public health. **The Lancet**, v. 377, n. 9765, p. 537–539, fev. 2011. DOI 10.1016/S0140-6736(10)62234-9. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0140673610622349>. Acesso em: 15 set. 2020.

WEITZEL, S. As novas configurações do Acesso Aberto: desafios e propostas. **Revista Eletrônica de Comunicação, Informação e Inovação em Saúde**, v. 8, n. 2, 2014. DOI 10.3395/receis.v8i2.447. Disponível em: <https://www.receis.iciet.fiocruz.br/index.php/receis/article/view/447>. Acesso em: 9 set. 2020.

WILKINSON, M. Datacite: The International Data Citation Initiative: Datasets Programme. **SSRN Electronic Journal**, 2010. DOI 10.2139/ssrn.1743431. Disponível em: <http://www.ssrn.com/abstract=1743431>. Acesso em: 25 set. 2020.

WILKINSON, M. D.; DUMONTIER, M.; AALBERSBERG, Ij. J.; APPLETON, G.; AXTON, M.; BAAK, A.; BLOMBERG, N.; BOITEN, J.-W.; DA SILVA SANTOS, L. B.; BOURNE, P. E.; BOUWMAN, J.; BROOKES, A. J.; CLARK, T.; CROSAS, M.; DILLO, I.; DUMON, O.; EDMUNDS, S.; EVELO, C. T.; FINKERS, R.; ... MONS, B. The FAIR Guiding Principles for scientific data management and stewardship. **Scientific Data**, v. 3, n. 1, p. 160018, dez. 2016. DOI 10.1038/sdata.2016.18. Disponível em: <http://www.nature.com/articles/sdata201618>. Acesso em: 9 set. 2020.

WOOD, C. C. As razões porque o Crossref existe e persiste [online]. **SciELO em Perspectiva**, 2018. Disponível em: <https://blog.scielo.org/blog/2018/07/17/as-razoes-porque-o-crossref-existe-e-persiste/#.X11ZI3lKiUk>. Acesso em: 9 set. 2020.

ZANETTI, D. A cultura do compartilhamento e a reprodutibilidade dos conteúdos. **C-Legenda - Revista do Programa de Pós-graduação em Cinema e Audiovisual**, v. 0, n. 25, 2012. Disponível em: <https://periodicos.uff.br/ciberlegenda/article/view/36884>.

ZENG, M. L.; QIN, J. **Metadata**. 2. ed. Chicago: Neal-Schuman, 2016.

ZIMAN, J. M. Information, Communication, Knowledge. **Nature**, v. 224, n. 5217, p. 318–324, out. 1969. DOI 10.1038/224318a0. Disponível em: <http://www.nature.com/articles/224318a0>. Acesso em: 9 set. 2020.

ZIMAN, John M. **An Introduction to Science Studies: The Philosophical and Social Aspects of Science and Technology**. 1. ed. [S. l.]: Cambridge University Press, 1984. DOI 10.1017/CBO9780511608360. Disponível em: <https://www.cambridge.org/core/product/identifier/9780511608360/type/book>. Acesso em: 9 set. 2020.