



UNIVERSIDADE FEDERAL DE SÃO CARLOS
CENTRO DE EDUCAÇÃO E CIÊNCIAS HUMANAS
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO

WENDY KETLEN PINTO DA SILVA

**ATRIBUIÇÃO DE PALAVRAS-CHAVE NA PRÁTICA DO
AUTOARQUIVAMENTO: estudo das diretrizes de repositórios de
dados de pesquisa ibero-americanos**

São Carlos – SP

2022

WENDY KETLEN PINTO DA SILVA

**ATRIBUIÇÃO DE PALAVRAS-CHAVE NA PRÁTICA DO
AUTOARQUIVAMENTO: estudo das diretrizes de repositórios de
dados de pesquisa ibero-americanos**

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Informação da Universidade Federal de São Carlos (UFSCar), como parte das exigências para a obtenção do título de Mestre em Ciência da Informação.

Área de concentração: Conhecimento, Tecnologia e Inovação.

Linha de pesquisa: Tecnologia, Informação e Representação.

Orientadora: Profa. Dra. Paula Regina Dal'Evedove

São Carlos – SP

2022

Silva, Wendy Ketlen Pinto da

Atribuição de palavras-chave na prática do autoarquivamento: estudo das diretrizes de repositórios de dados de pesquisa ibero-americanos / Wendy Ketlen Pinto da Silva -- 2022.
93f.

Dissertação (Mestrado) - Universidade Federal de São Carlos, campus São Carlos, São Carlos
Orientador (a): Paula Regina Dal'Evedove
Banca Examinadora: Paula Regina Dal'Evedove, Ana Lúcia Terra, Flavia Maria Bastos, Zaira Regina Zafalon
Bibliografia

1. Palavras-chave. 2. Autoarquivamento. 3. Repositórios de dados de pesquisa. I. Silva, Wendy Ketlen Pinto da. II. Título.

Ficha catalográfica desenvolvida pela Secretaria Geral de Informática (SIn)

DADOS FORNECIDOS PELO AUTOR

Bibliotecário responsável: Ronildo Santos Prado - CRB/8 7325



UNIVERSIDADE FEDERAL DE SÃO CARLOS

Centro de Educação e Ciências Humanas
Programa de Pós-Graduação em Ciência da Informação

Folha de Aprovação

Defesa de Dissertação de Mestrado da candidata Wendy Ketlen Pinto da Silva, realizada em 11/05/2022.

Comissão Julgadora:

Profa. Dra. Paula Regina Da'Evedove (UFSCar)

Profa. Dra. Zaira Regina Zafalon (UFSCar)

Profa. Dra. Ana Lúcia Silva Terra (UC)

Profa. Dra. Flavia Maria Bastos (UNESP)

À minha mãe, que foi a primeira a incentivar meus estudos desde tenra idade, buscando por oportunidades que, às vezes, ela nem mesmo sabia que existiam.

AGRADECIMENTOS

Gostaria de iniciar agradecendo à minha orientadora, prof.^a Paula, por ter me acompanhado nessa caminhada, me auxiliando imensamente durante todo o processo, me acolhendo quando cheguei à UFSCar e me incentivado a seguir.

À todos os professores que fazem parte do PPGCI-UFSCar e que contribuíram, de alguma forma, durante as aulas e eventos, sempre buscando transmitir conhecimento a quem estivesse disposto a ouvir. Saibam que as aulas foram preciosas para mim.

Ao meu marido, Felipe, por ter me auxiliado em diversos momentos durante todo o processo da pesquisa, por ter ficado tão animado quanto eu, mas principalmente por ser meu maior incentivador nessa caminhada pela vida acadêmica que pode ser, por vezes, desafiadora.

RESUMO

Os repositórios de dados de pesquisa surgem como ambientes informacionais digitais comprometidos com o armazenamento, gestão, acesso e preservação dos registros provenientes das pesquisas científicas, cada vez mais estimulados pelo desenvolvimento da Ciência Aberta. Apesar da ampla discussão envolvendo dados de pesquisa, questiona-se como as diretrizes disponibilizadas amparam o autor-indexador no preenchimento do metadado assunto durante o processo de autoarquivamento em repositórios de dados de pesquisa? Diante disso, o objetivo geral da pesquisa é analisar as recomendações para a atribuição de palavras-chave pelo autor-indexador em diretrizes de repositórios de dados de pesquisa de países integrantes da ibero-américa que adotam o autoarquivamento, no intuito de contribuir para o desenvolvimento de estudos teóricos e operacionais dedicados ao metadado assunto. Para tanto, desenvolve-se um estudo exploratório e descritivo com abordagem qualitativa que alia pesquisa bibliográfica e documental, esta última a partir da identificação e seleção de repositórios de dados de pesquisa ibero-americanos que adotam o autoarquivamento, obtido da ferramenta Re3data. Foram identificados 26 repositórios de dados de pesquisa, cujas diretrizes e respectivos ambientes digitais foram analisados por meio de seis indicadores, visando à construção de considerações analíticas e comparativas. Os resultados apontam para a falta de instruções para a atribuição de palavras-chave, mesmo quando os repositórios de dados de pesquisa possuem diretrizes formalizadas, além da falta de menção ao metadado assunto e palavras-chave, assim como instruções claras para a representação dos conteúdos dos dados de pesquisa pelo autor-pesquisador. A maior parte dos repositórios de dados de pesquisa recomendam o uso de linguagens documentárias, assim como realizam a validação dos termos empregados pelos autores-pesquisadores. Como conclusões, tem-se um cenário preocupante sobre a atribuição de palavras-chave na prática do autoarquivamento em repositórios de dados de pesquisa ibero-americanos, sendo oportunos estudos complementares que proponham orientações referentes ao metadado assunto nesses sistemas de informação.

Palavras-chave: Palavras-chaves. Termos de indexação. Autoarquivamento. Repositórios de dados de pesquisa. Ibero-américa.

ABSTRACT

Research data repositories emerge as digital informational environments committed to the storage, management, access, and preservation of records from scientific research, increasingly stimulated by the development of Open Science. Despite the broad discussion involving research data, one wonders how the available guidelines support the author-indexer in filling in the subject metadata during the self-archiving process in research data repositories? Given this, the general objective of the research is to analyze the recommendations for the attribution of keywords by the author-indexer in guidelines of research data repositories of countries belonging to Ibero America that adopt self-archiving to contribute to the development of theoretical and operational studies dedicated to the subject metadata. Therefore, an exploratory and descriptive study with a qualitative approach is developed that combines bibliographic and documentary research. The latter is based on identifying and selecting repositories of Iberoamerican research data that adopt self-archiving, obtained from the Re3data tool. Twenty-six research data repositories were identified, whose guidelines and respective digital environments were analyzed using six indicators to construct analytical and comparative considerations. The results point to the lack of instructions for the attribution of keywords, even when the research data repositories have formalized guidelines, in addition to the lack of mention of the subject metadata and keywords, as well as clear instructions for the representation of the contents of the research data by the author-researcher. Most research data repositories recommend the use of documentary languages and validate the terms used by the authors-researchers. In conclusion, there is a worrying scenario regarding the attribution of keywords in self-archiving in Iberoamerican research data repositories, and complementary studies are opportune to propose guidelines regarding the subject metadata in these information systems.

Keywords: *Keywords. Indexing Terms. Self-archiving. Research data repositories. Ibero America.*

LISTA DE ILUSTRAÇÕES

Gráfico 1 – Porcentagem de repositórios que adotam o autoarquivamento.....	64
Gráfico 2 – Padrões de metadados utilizados.....	65
Gráfico 3 – Diretrizes formais estabelecidas.....	68
Gráfico 4 – Indicação do metadado assunto.....	69
Gráfico 5 – Indicação de palavras-chave.....	71
Gráfico 6 – Instruções para o preenchimento do metadado assunto/palavras-chave.	73
Gráfico 7 – Indicação de linguagens documentárias.....	75
Gráfico 8 – Validação profissional.....	77

LISTA DE TABELAS

Tabela 1 – Repositórios de dados de pesquisa da ibero-américa que adotam o autoarquivamento.....	24
Tabela 2 – Indicadores para avaliação das diretrizes.	26
Tabela 3 – Os princípios FAIR.	51

LISTA DE ABREVIATURAS E SIGLAS

ABCD	<i>Access to Biological Collection Data</i>
BDTD	Biblioteca Digital Brasileira de Teses e Dissertações
BD2K	<i>Big Data to Knowledge</i>
BOAI	<i>Budapest Open Access Initiative</i>
BRAPCI	Base de Dados Referencial de Artigos de Periódicos em Ciência da Informação
BSC	<i>Barcelona Supercomputer Center</i>
CDDC	Curadoria Digital de Dados Científicos
CIAT	<i>International Center for Tropical Agriculture Dataverse</i>
CIMMYT	Centro Internacional de Melhoramento de Milho e Trigo
CNPq	Conselho Nacional de Desenvolvimento Científico e Tecnológico
CSEE	Ciência da Computação e Engenharia Elétrica
DCMI	<i>Dublin Core Metadata Initiative</i>
DDI	<i>Data Documentation Initiative</i>
DTL	<i>Dutch Techcentre for Life Sciences</i>
DOI	<i>Digital Object Identifier</i>
DwC	<i>Darwin Core</i>
EML	<i>Ecological Metadata Language</i>
EOSC	<i>Commission High Level Expert Group on the European Open Science Cloud</i>
FAIR	<i>Findable, Accessible, Interoperable, Reusable</i>
ForestGEO	<i>Forest Global Earth Observatory</i>
GES	Geografia e Sistemas Ambientais
GNU	<i>General Public License</i>
HP	<i>Hewlett Packard</i>

IBICT	Instituto Brasileiro de Informação em Ciência e Tecnologia
IS	Sistemas de Informação
LISA	<i>Library & Information Science Abstracts</i>
MIT	<i>Massachusetts Institute of Technology</i>
NIH	<i>American National Institutes for Health</i>
OAI	<i>Open Archives Initiative</i>
OAI-ORE	<i>Open Archives Initiative Object Reuse and Exchange</i>
OAI-PMH	<i>Open Archives Initiative for Metadata Harvesting</i>
PID	<i>Persistent Identifier</i>
PMH	<i>Protocol for Metadata Harvesting</i>
PPBio	Programa de Pesquisa em Biodiversidade
IODP	Programa Internacional de Descoberta do Oceano
Re3data.org	<i>Registry of Research Data Repositories</i>
SciELO	<i>Scientific Electronic Library Online</i>
SISIB	Diretoria de Serviços de Informação e Bibliotecas
SNDB	Sistema Nacional de Dados Biológicos
TICs	Tecnologias de Informação e Comunicação
UAB	<i>Universitat Autònoma de Barcelona</i>
UFSCar	Universidade Federal de São Carlos
UNAM	<i>Universidad Nacional Autónoma de México</i>
UNED	Universidade Nacional de Educação à Distância
UNESP	Universidade Estadual Paulista
XML	<i>Extensible Markup Languageem</i>

SUMÁRIO

1	INTRODUÇÃO	14
1.1	OBJETIVOS	17
1.2	JUSTIFICATIVA	18
1.3	ESTRUTURA DA PESQUISA	19
2	PROCEDIMENTOS METODOLÓGICOS	21
2.1	UNIVERSO DA PESQUISA	26
3	REFERENCIAL TEÓRICO	38
3.1	O MOVIMENTO DE ACESSO ABERTO	38
3.2	REPOSITÓRIOS DE DADOS DE PESQUISA.....	43
3.3	DADOS DE PESQUISA.....	47
3.4	OS PRINCÍPIOS <i>FAIR</i> EM REPOSITÓRIOS DE DADOS DE PESQUISA	50
3.5	ATRIBUIÇÃO DE PALAVRAS-CHAVE COMO VALOR DO METADADO ASSUNTO.....	56
4	ANÁLISE E DISCUSSÃO DOS RESULTADOS	63
4.1	DIRETRIZES.....	67
4.2	METADADO ASSUNTO	69
4.3	PALAVRAS-CHAVE.....	70
4.4	PREENCHIMENTO DO METADADO ASSUNTO/PALAVRAS-CHAVE.....	71
4.5	LINGUAGEM DOCUMENTÁRIA	73
4.6	VALIDAÇÃO DAS PALAVRAS-CHAVE POR PROFISSIONAIS	76
5	CONSIDERAÇÕES FINAIS	79
	REFERÊNCIAS	83
	ANEXO A – Identificação dos padrões de metadados por repositório de dados de pesquisa	90
	ANEXO B – Diretrizes e demais documentos utilizados para análise	92

1 INTRODUÇÃO

Este estudo, desenvolvido no âmbito da Linha de pesquisa “Tecnologia, Informação e Representação” do Programa de Pós-Graduação em Ciência da Informação da Universidade Federal de São Carlos (UFSCar), possui como tema de pesquisa o metadado assunto em repositórios de dados de pesquisa ibero-americanos, cuja delimitação consiste na análise de diretrizes para a atribuição de palavras-chave por autores-indexadores na prática do autoarquivamento.

O surgimento da Internet e os avanços nas Tecnologias de Informação e Comunicação viabilizaram grandes mudanças na Sociedade da Informação, especialmente na forma como esta se comunica. Isso inclui a comunicação científica e os serviços que atuam com informação ao redor do mundo. A expansão da comunicação em rede e o aumento do fluxo informacional possibilitou a ampliação na transmissão de informações, o que impactou a produção e o compartilhamento da produção científica, que encontrava novos meios para ser disponibilizada através das tecnologias emergentes.

Em meio a esses novos formatos de compartilhamento de informações científicas voltados para o ambiente digital a Ciência Aberta se manifestou, trazendo consigo novas estratégias e ferramentas em defesa de uma comunicação científica mais abrangente e irrestrita. Alicerçado nisso, o Movimento de Acesso Aberto (*open access*) teve seu início a partir de várias iniciativas, dentre as quais encontra-se a Declaração de Budapeste¹, Declaração de Bethesda² e a Declaração de Berlim³, consideradas as mais significativas para que o acesso aberto se tornasse uma realidade em todo o mundo.

O autoarquivamento ou auto-depósito foi estabelecido pela *Budapest Open Access Initiative* como a primeira estratégia para se alcançar os objetivos de uma Ciência mais aberta, uma vez que essa prática pode partir do próprio autor-pesquisador ou do mediador (profissional ou gerenciador do sistema) para a divulgação de pesquisas e dados científicos, tornando-os disponíveis para acesso e recuperação por qualquer usuário de qualquer parte do mundo, de forma rápida e gratuita. Em vista disso, os dados científicos se tornaram um dos ativos mais

¹ *Budapest Open Access Initiative*, de 2002.

² *Bethesda Statement on Open Access Publishing*, de 2003.

³ *Berlin Declaration on Open Access to Knowledge in Science & Humanities*, de 2003.

significativos para a comunidade científica, de forma que não somente a versão final da pesquisa fosse compartilhada, mas também todos os materiais envolvidos em seu processo.

Nesse contexto, os repositórios de dados de pesquisa se caracterizam como a principal ferramenta disponível para o gerenciamento e compartilhamento de dados provenientes de pesquisas científicas, auxiliando no crescimento dos conteúdos digitais produzidos por universidades e instituições. Os dados de pesquisa são depositados nos repositórios pelo próprio autor-pesquisador ou por um profissional que faz a mediação entre o autor-pesquisador e o sistema de informação. A etapa do autoarquivamento no sistema de informação deve ser coordenada por uma série de diretrizes, entre as quais estão orientações específicas voltadas para a representação do recurso informacional a ser disponibilizado para a comunidade. Uma vez que os termos/palavras-chave empregados no ato do autoarquivamento operam como pontos de acesso, manifesta-se a preocupação com o preenchimento do metadado assunto nesses ambientes digitais colaborativos.

Cabe destacar que durante o autoarquivamento o autor age como indexador especialista do domínio que não tem conhecimento de que a atribuição de palavras-chave é um processo de indexação de assuntos, o qual precisa ser orientado mediante o estabelecimento de diretrizes específicas. Falta de procedimentos adequados para a representação do conteúdo de dados de pesquisa pelo autor-indexador impacta no compartilhamento e acesso desses recursos informacionais, inviabilizando a reprodutibilidade científica.

A pertinência de pesquisas dedicadas ao metadado assunto em repositórios de dados de pesquisa decorre do aumento da prática do autoarquivamento para a representação do conteúdo de dados de pesquisa e a pouca literatura dedicada à questão. No contexto da Ciência da Informação brasileira, Freitas (2019) investigou a consistência na indexação livre realizada pelo autor-pesquisador no processo do autoarquivamento em comparação com a de um bibliotecário indexador. Nesta investigação, a partir de uma abordagem investigativa centrada na avaliação da indexação de teses e dissertações do Repositório Institucional da Universidade Federal de São Carlos (UFSCar), a autora conclui que o autor-pesquisador apresenta uma indexação mais exaustiva em comparação com a indexação praticada pelo bibliotecário indexador. Diante da falta de precisão dos termos/palavras-chave atribuídos pelos autores/pesquisadores na descrição do

metadado assunto de teses e dissertações, o estudo sugere que o sistema de informação disponibilize linguagens documentárias para acesso e uso pelos autores/pesquisadores, além da validação do metadado assunto pelos profissionais bibliotecários do repositório institucional e elaboração e disponibilização de uma política de indexação.

O estudo conduzido por Viana e Dal'Evedove (2021) investigou a indexação de dados de pesquisa no âmbito da Rede de Repositórios de Dados Científicos do Estado de São Paulo, a partir das políticas de informação estabelecidas por cada instituição integrante. Dentre os resultados obtidos, as autoras constataam a incipiência de diretrizes específicas para o autodepósito nos repositórios de dados de pesquisa vinculados à Rede, sobretudo, na atribuição de palavras-chave para o preenchimento do metadado assunto. Os resultados obtidos nesta pesquisa contribuem com os estudos de representação por assunto em repositórios digitais, sendo oportuna a condução de pesquisas complementares sobre a temática.

Em Freitas, Dal'Evedove e Tartarotti (2021), as autoras analisaram a presença de políticas de autoarquivamento em repositórios institucionais brasileiros, com destaque para as regras estabelecidas para a representação de assunto. Como resultado, o estudo revelou que dos 108 repositórios institucionais brasileiros listados no Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT), apenas sete possuem políticas de autoarquivamento, sendo que em apenas uma das políticas analisadas existe indicações sobre a qualidade dos metadados, porém sem pontuar recomendações para o preenchimento do metadado assunto.

Questões envolvendo o metadado assunto, com destaque para pesquisas sobre indexação de assunto em repositórios digitais (SANTOS, 2017a; SANTOS; NEVES, 2018; TARTAROTTI, 2019; TARTAROTTI; DAL'EVEDOVE; FUJITA, 2020; FREITAS; DAL'EVEDOVE; TARTAROTTI, 2021, dentre outros) e atribuição de palavras-chave pelo autor-pesquisador em ambientes digitais de comunicação científica aberta (MIGUÉIS et. al., 2013; SANTOS, 2017b; OLIVEIRA et.al, 2020; FUJITA; TARTAROTTI, 2020; TERRA et. al; 2021, dentre outros) é uma realidade que corrobora para a importância do desenvolvimento de pesquisas nessa direção. Contribuir com discussões nessa direção mostra-se oportuno, especialmente quando dados de pesquisa e repositórios de dados de pesquisa se projetam como temas atuais e em franco desenvolvimento na literatura especializada.

Diante desse cenário, mostra-se oportuno a condução de estudos mais

aprofundados sobre o processo de atribuição de palavras-chave no momento do preenchimento do metadado assunto em repositórios de dados de pesquisa que adotam o autoarquivamento. Com isso, questiona-se: como as diretrizes estabelecidas amparam o autor-indexador no preenchimento do metadado assunto durante o processo de autoarquivamento em repositórios de dados de pesquisa?

Portanto, o enfoque desta pesquisa recai nas diretrizes dos repositórios de dados de pesquisa ibero-americanos, com especial atenção para a presença de diretrizes que ofereçam orientações e recomendações específicas para o autor-indexador na atribuição de palavras-chave.

Nota-se a emergência de discussões nessa direção em razão do pouco envolvimento dos pesquisadores quando o problema abordado é a representação do conteúdo de dados de pesquisa durante o preenchimento do metadado assunto, sendo oportuna a condução de uma pesquisa que identifique e discuta esta prática em repositórios de dados de pesquisa.

Tencionando responder à questão levantada, parte-se para a exposição dos objetivos geral e específicos da pesquisa.

1.1 OBJETIVOS

O objetivo geral da pesquisa é analisar as recomendações para a atribuição de palavras-chave pelo autor-indexador em diretrizes de repositórios de dados de pesquisa de países integrantes da ibero-américa que adotam o autoarquivamento, no intuito de contribuir para o desenvolvimento de estudos teóricos e operacionais dedicados ao metadado assunto.

A fim de alcançar o objetivo geral, foram delineados os seguintes objetivos específicos:

1. Conceituar os repositórios de dados de pesquisa pela perspectiva da Ciência da Informação e da Ciência Aberta, com foco no metadado assunto e na prática do autoarquivamento;
2. Identificar os repositórios de dados de pesquisa ibero-americanos que adotam o autoarquivamento e suas respectivas diretrizes;
3. Caracterizar a atribuição de palavras-chave a partir das diretrizes estabelecidas pelos repositórios de dados de pesquisa ibero-americanos que

- adotam o autoarquivamento; e
4. Evidenciar a situação atual das diretrizes quanto ao preenchimento do metadado assunto.

1.2 JUSTIFICATIVA

A nova realidade vivenciada na representação de conteúdos de documentos, sobretudo, a partir do processo de autoarquivamento inserido no contexto dos repositórios digitais, convoca pesquisadores e profissionais na reflexão e construções de elementos teóricos e operacionais que subsidiem o processo praticado pelo autor-indexador, salvaguardando uma descrição e recuperação por assuntos que seja verdadeiramente relevante para outros usuários.

Discussões nesta direção são objetos de interesse do Grupo de Pesquisa Organização do Conhecimento e Humanidades Digitais⁴ da Universidade Federal de São Carlos, com especial atenção aos problemas oriundos do ambiente digital na atividade de indexação de assuntos. Por esta razão, a presente pesquisa busca contribuir com o entendimento do processo de representação de conteúdos praticado em repositórios de dados de pesquisa que adotam o autoarquivamento.

Espera-se que os resultados aqui identificados colaborem, em particular: a) para o desenvolvimento de orientações para diretrizes disponibilizadas por repositórios de dados de pesquisa brasileiros; b) para conscientizar sobre a importância do metadado assunto na representação de recursos informacionais digitais, sobretudo provenientes de pesquisas científicas; e c) para apoiar o Movimento de Ciência Aberta ao apresentar considerações acerca da atual situação do processo de atribuição de palavras-chave na prática do autoarquivamento em repositórios de dados de pesquisa ibero-americanos, com intuito de contribuir para o desenvolvimento de estudos acerca do metadado assunto e, por consequência, do acesso e recuperação por assunto dos dados de pesquisa.

Busca-se, com isso, oferecer à comunidade científica da Ciência da Informação subsídios para o avanço das discussões na temática. Neste sentido, espera-se que os resultados obtidos com a pesquisa contribuam para o

⁴ Cadastrado no Diretório dos Grupos de Pesquisa do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), com acesso às informações complementares pelo link: <http://dgp.cnpq.br/dgp/espelhogrupo/340090>.

desenvolvimento do tema e incentivem outros pesquisadores a trabalharem com este eixo investigativo.

Entende-se que o desenvolvimento da presente pesquisa colabora com os estudos na área de Organização do Conhecimento dedicados à representação por assunto em ambientes digitais, com destaque para resultados que favoreçam o acesso e a recuperação de dados de pesquisa em sistemas híbridos de informação que adotam o autoarquivamento a partir da especificidade no preenchimento do metadado assunto, contribuindo com a celeridade do processo de difusão de publicações científicas. Igualmente, seus resultados podem fornecer *insights* sobre as práticas de metadados de assunto que estão em vigor ou em discussão em repositórios de dados de pesquisa de países ibero-americanos, considerando a relevância desses sistemas de informação no apoio ao desenvolvimento e avanço das pesquisas científicas.

A escolha pela condução de uma pesquisa que considere os repositórios de dados de pesquisa dos países pertencentes à comunidade ibero-americana se justifica em razão do número pequeno desse tipo de repositório no contexto brasileiro, além da facilidade da pesquisadora com os idiomas português e espanhol, línguas predominantes nesses países. Além disso, a condução de um estudo analítico amplo, como é o caso nesta pesquisa, favorece análises comparativas das principais recomendações adotadas por diferentes repositórios de dados de pesquisa, cujos achados podem colaborar com o desenvolvimento de futuras pesquisas dedicadas à elaboração de diretrizes para a atribuição de palavras-chave no momento do preenchimento do metadado assunto nesses sistemas de informação.

1.3 ESTRUTURA DA PESQUISA

A Dissertação está estruturada em cinco seções. Esta seção inicial é dedicada à exposição do tema e do problema de pesquisa, assim como apresenta o objetivo geral, objetivos específicos e a justificativa para a sua vinculação e desenvolvimento na Ciência da Informação.

A segunda seção apresenta os procedimentos metodológicos que amparam o seu desenvolvimento. Neste ponto, descreve-se a etapa da pesquisa bibliográfica que fundamenta a construção de considerações teóricas, assim como a forma de

identificação dos repositórios de dados de pesquisa que adotam o autoarquivamento, tendo como universo de pesquisa os países ibero-americanos. Do mesmo modo, é exposto como as análises das diretrizes dos repositórios selecionados foram realizadas.

A terceira seção é dedicada ao referencial teórico, sendo apresentadas considerações sobre o Movimento de Acesso Aberto e as iniciativas que o impulsionaram, a prática do autoarquivamento, repositórios de dados de pesquisa, dados de pesquisa, princípios *FAIR* e a maneira que podem ser aplicados em repositórios de dados de pesquisa, representação temática da informação em ambientes digitais, com foco para a indexação de assunto, palavras-chave e o metadado assunto. Busca-se, com isso, apresentar uma revisão teórica que contemple as principais recomendações envolvendo o metadado assunto em repositórios digitais, com perspectivas de vinculação da discussão para o cenário específico dos repositórios de dados de pesquisa, resultando em uma sintetização que contribua para o desenvolvimento do tema na Ciência da Informação brasileira.

A quarta seção contempla as análises das diretrizes dos repositórios de dados de pesquisa ibero-americanos selecionados. Apresenta e discute os achados da pesquisa, com menção sobre como os resultados se relacionam com as considerações advindas da literatura especializada, além de discorrer sobre a situação atual das diretrizes quanto ao preenchimento do metadado assunto.

Por fim, a quinta e última seção discorre sobre as conclusões da pesquisa, com recomendações para o desenvolvimento de estudos futuros.

2 PROCEDIMENTOS METODOLÓGICOS

Esta pesquisa caracteriza-se como exploratória e descritiva com abordagem qualitativa, englobando pesquisa bibliográfica e documental. A pesquisa se caracteriza ainda por sua natureza aplicada, uma vez que visa oferecer subsídios para a elaboração de diretrizes que amparem o autor-indexador no momento da atribuição de palavras-chave na prática do autoarquivamento em repositórios de dados de pesquisa.

A primeira etapa contemplou a pesquisa bibliográfica, a qual viabilizou a construção de considerações analíticas e reflexivas, além de oportunizar a análise e discussão dos dados obtidos com a pesquisa documental. A revisão de literatura contemplou diversas fontes de informação dedicadas ao Movimento de Acesso Aberto, autoarquivamento, repositórios de dados de pesquisa, dados de pesquisa, representação temática da informação, metadado assunto, palavras-chave, além de outros temas correlatos.

O levantamento bibliográfico foi pautado na busca e coleta de publicações científicas disponíveis em bases de dados nacionais como a Base de Dados Referencial de Artigos de Periódicos em Ciência da Informação (BRAPCI) e a Biblioteca Digital Brasileira de Teses e Dissertações (BDTD); assim como bases de dados internacionais disponíveis para consulta a partir do Portal de Periódicos CAPES⁵, tais como *Library & Information Science Abstracts* (LISA), *Scientific Electronic Library Online* (Scielo), *Scopus* e *Web of Science*. Os temas centrais da pesquisa foram empregados como termos no processo de busca, sendo as estratégias de busca aplicadas sem delimitação temporal, a fim de identificar o maior número possível de estudos nacionais e internacionais e apresentar, com isso, o desenvolvimento e evolução da discussão envolvendo o escopo da pesquisa.

Para tanto, foram empregados termos e respectivos sinônimos em português, inglês e espanhol, elegíveis a partir de consulta ao Tesouro Brasileiro de Ciência da Informação⁶, quais sejam: repositório de dados; repositório de pesquisa; repositório de dados de pesquisa; repositório de dados científicos; dados de pesquisa; dados científicos; autoarquivamento; auto-depósito; representação temática; representação

⁵ Criado e mantido pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES).

⁶ Disponível em: <http://sitehistorico.ibict.br/publicacoes-e-institucionais/tesauro-brasileiro-de-ciencia-da-informacao-1/tesauro-brasileiro-de-ciencia-da-informacao>

por assunto; indexação; indexação de assuntos; palavras-chave; termos de indexação; cabeçalhos de assunto; descritores; metadado de assunto; metadado temático; metadados de acesso; política de informação; e política de indexação.

Com isso, as estratégias de busca⁷ empregadas são descritas a seguir:

- a) “repositório de dados” OR “repositório de pesquisa” OR “repositório de dados de pesquisa” OR “repositório de dados científicos” AND “dados de pesquisa” OR “dados científicos”;
- b) “repositório de dados” OR “repositório de pesquisa” OR “repositório de dados de pesquisa” OR “repositório de dados científicos” AND “autoarquivamento” OR “auto-depósito”;
- c) “repositório de dados” OR “repositório de pesquisa” OR “repositório de dados de pesquisa” OR “repositório de dados científicos” AND “indexação” OR “indexação de assuntos” OR “representação temática” OR “representação por assunto” OR “política de indexação” OR “política de informação” OR “metadado de assunto” OR “metadado temático” OR “metadados de acesso”;
- d) “repositório de dados” OR “repositório de pesquisa” OR “repositório de dados de pesquisa” OR “repositório de dados científicos” AND “palavras-chave” OR “termos de indexação” OR “cabeçalhos de assunto” OR “descritores”; e
- e) “autoarquivamento” OR “auto-depósito” AND “indexação” OR “indexação de assuntos” OR “representação temática” OR “representação por assunto” OR “política de indexação” OR “política de informação” OR “metadado de assunto” OR “metadado temático” OR “metadados de acesso”.

Como critérios de inclusão dos materiais para a leitura e interpretação nesta etapa de pesquisa bibliográfica, foram selecionadas fontes bibliográficas em formato digital⁸ nos idiomas português, inglês e espanhol, a partir da busca por título, palavras-chave e resumo nas referidas bases de dados, desde que disponibilizadas

⁷ Ressalta-se que essas mesmas estratégias de busca foram empregadas com os termos em inglês e espanhol em cada uma das Bases de Dados selecionadas na pesquisa.

⁸ Diante da atual pandemia de COVID-19 vivenciada na atualidade, não foi possível obter acesso aos materiais impressos como livros e periódicos científicos disponíveis na Biblioteca Comunitária da Universidade Federal de São Carlos em razão do seu fechamento, conforme a Portaria do GR nº 4370, de 14 de março de 2020, que determinou a suspensão de aulas e atividades curriculares a partir de 16 de março; assim como da Portaria GR nº 4380, de 20 de março de 2020, que prorrogou a suspensão de aulas, atividades Curriculares e administrativas por tempo indeterminado.

em formato completo e circunscritos no contexto da pesquisa. Portanto, o interesse maior foi por materiais dedicados à discussão da atribuição de palavras-chave, com foco na prática do autoarquivamento, tendo como base outras investigações mais amplas sobre o metadado assunto, repositórios digitais, indexação de assunto e representação por assunto em ambientes digitais.

A etapa da pesquisa documental foi conduzida a partir da identificação e seleção do universo da pesquisa, composto pelos repositórios de dados de pesquisa de países ibero-americanos que adotam o autoarquivamento. Para a identificação dos repositórios de dados de pesquisa, recorreu-se ao diretório *Registry of Research Data Repositories* (Re3data.org) enquanto recurso que fornece informações abrangentes sobre esses sistemas de informação.

Com mais de 2.600 repositórios registrados, o Re3data é um registro global de repositórios que cobre repositórios de dados de pesquisa de diferentes disciplinas acadêmicas. Inclui repositórios de dados de pesquisa que permitem o armazenamento permanente e o acesso a conjuntos de dados para pesquisadores, órgãos de financiamento, editores e instituições acadêmicas. Surgiu no outono de 2012 e foi financiado pela Fundação Alemã de Pesquisa (DFG).

Editores e periódicos como *Copernicus Publications*, *PeerJ*, *Springer* e *Nature's Scientific Data* referem-se ao Re3data em suas políticas editoriais como uma ferramenta para a fácil identificação de repositórios apropriados para armazenar dados de pesquisa. O uso do Re3data também é recomendado nas “Diretrizes sobre Acesso Aberto a Publicações Científicas e Dados de Pesquisa no Horizonte 2020” da Comissão Europeia.

Nesse âmbito, os repositórios de dados de pesquisa em países ibero-americanos que adotam o autoarquivamento e selecionados a partir do Re3data somam o total de 26, sendo listados na Tabela 1.

A identificação dos repositórios de dados de pesquisa foi realizada no mês de outubro de 2021, sendo atualizada em fevereiro de 2022. Vale ressaltar que, apesar de ser um diretório para repositórios, bases de dados também estão registradas no Re3data, as quais não foram consideradas para fins de inclusão. Da mesma forma, os repositórios exclusivamente institucionais não foram contabilizados, pois entende-se que não são o foco da pesquisa.

Tabela 1 – Repositórios de dados de pesquisa da ibero-américa que adotam o autoarquivamento.

País	Repositórios de dados de pesquisa por país	Repositórios de dados de pesquisa que adotam autoarquivamento	Nomes dos repositórios que adotam autoarquivamento	Repositórios com acesso indisponível
Argentina	6	4	Produção acadêmica UCC (híbrido)	Portal de Datos Genómicos del SNDG Portal de datos del mar - SNDM
			Portal de datos de Biodiversidad	
			SEDICI (híbrido)	
			Humadoc repositorio (híbrido)	
Brasil	12	7	Globe	Base de Datos Científicos da Universidade Federal do Paraná
			PPBio Data Repository	
			IBICT Cariniana Dataverse Network	
			Scielo Data	
			CEDAP Reasearch data repository	
			Repositório Institucional UNESP (híbrido)	
Repositório de dados de pesquisa do Instituto Federal Goiano - campus Urutaí				
Chile	2	1	Repositorio de datos de investigación de la Universidad de Chile	
Colômbia	6	2	CIAT Dataverse	
			Repositorio de datos de investigación de la universidad del Rosario	
Espanha	20	6	IoChem – BD	SIOR - Social Impact of Science
			UPCommons - Research Data	
			Universitat Obeta de Catalunya O2 Dades (híbrido)	
			Digibug (híbrido)	Dipòsit Digital de Documents de la Universitat Autònoma de Barcelona
			RiuNet (híbrido)	
Repositório Digital de Documentos UAB (híbrido)				
México	8	3	CIMMYT - Research data and software repository network	
			clRcle	
			RU-IIEc (híbrido)	
Panamá	2	0		
Peru	3	0		
Portugal	9	3	DataRepositorium - Repositório de dados da universidade do Minho	
			KiMoSys - Kinetic models of biological systems	
			DadosIPB	
	TOTAL	TOTAL		TOTAL
	68	26		5

Fonte: Dados de pesquisa.

Alguns repositórios de dados de pesquisa não se encontravam disponíveis até o momento deste estudo, são eles: *Portal de Datos Genómicos del SNDG* e *Portal de datos del mar SNDM*, Argentina; Base de Datos Científicos da Universidade Federal do Paraná, Brasil; *Social Impact of Science*; e *Dipòsit Digital de Documents de la Universitat Autònoma de Barcelona*, Espanha. Sendo assim, é possível que o número total de repositórios de dados de pesquisa que adotam o autoarquivamento na ibero-américa seja maior em pesquisas futuras.

O restante dos países (Bolívia, Costa Rica, Cuba, Equador, El Salvador, Guatemala, Honduras, Nicarágua, Paraguai, Porto Rico, República Dominicana, Uruguai e Venezuela) não possuem até o momento repositórios de dados de pesquisa registrados no Re3data. Quanto ao Panamá e Peru, na atualização de fevereiro de 2022 os países aparecem com 2 e 3 repositórios registrados, respectivamente. Contudo, por serem repositórios de dados de pesquisa que não adotam o autoarquivamento, não entraram na pesquisa.

A partir da identificação e seleção dos repositórios de dados de pesquisa ibero-americanos, realizou-se uma análise exaustiva em cada um dos ambientes para a identificação de suas respectivas diretrizes e outros documentos que regulamentam e/ou orientam o processo de atribuição de palavras-chave dos dados de pesquisa incluídos no sistema, direcionadas para o autor-pesquisador ou mediador.

A análise foi realizada a partir da definição de questões que guiaram esse processo pelas diretrizes dos repositórios de dados de pesquisa selecionados. Para tanto, utilizou-se o *software* Excel para a tabulação dos dados e os gráficos foram gerados a partir de um *notebook* Python na versão 3.7.13, em ambiente *Google Colab*, e as bibliotecas:

- *notebook* (v5.3.1): ambiente de desenvolvimento;
- *matplotlib* (v.3.2.2): para visualização dos dados;
- *pandas* (v1.3.5): para a análise e manipulação dos dados;
- *seaborn* (v0.11.2): para estilização da visualização dos dados; e
- *numpy* (v1.21.5): para as operações matemáticas.

Os indicadores 2 e 4 foram previamente definidos, pois fazem parte do escopo principal da pesquisa, enquanto o restante dos indicadores surgiram a partir das observações acerca das necessidades contidas nas próprias diretrizes analisadas e

nos ambientes dos repositórios de dados de pesquisa. Por fim, os seguintes indicadores foram estabelecidos, conforme exibido na Tabela 2.

Tabela 2 – Indicadores para avaliação das diretrizes.

Indicador	Questão norteadora
Diretrizes	Há documento formal?
Metadado Assunto	Há menção ao metadado assunto?
Palavras-chave	Há menção às palavras-chave?
Preenchimento do metadado/palavra-chave	Há instruções quanto ao preenchimento do metadado assunto e/ou atribuição de palavras-chave?
Linguagem documentária	Há indicação para a utilização de linguagens documentárias?
Validação das palavras-chave por profissionais	Há curadoria dos metadados por profissionais?

Fonte: Elaborado pela autora.

Buscou-se, com isso, sistematizar as principais recomendações e descrever o cenário atual das diretrizes e orientações relativas à atribuição de palavras-chave pelo autor-indexador durante o autoarquivamento em repositórios de dados de pesquisa ibero-americanos.

Desse modo, foi possível conhecer o processo de representação do conteúdo de dados de pesquisa definido por cada um dos repositórios de dados de pesquisa analisados. Doravante, apresentar considerações analíticas e comparativas sobre a atual situação do metadado assunto em repositórios de dados de pesquisa na ibero-américa.

A seguir, são apresentados cada um dos repositórios de dados de pesquisa contemplados nesta pesquisa, como forma de contextualização do histórico e propósito de cada um dos sistemas de informação aqui investigados.

2.1 UNIVERSO DA PESQUISA

Os vinte e seis repositórios de dados de pesquisa mencionados na Tabela 1 são apresentadas nesta etapa, contemplando informações relevantes acerca da história, tipos de documentos, formatos de dados aceitos, entre outros elementos.

Argentina:

— Produção acadêmica UCC

É o repositório digital institucional da Universidade Católica de Córdoba. Conta com a produção acadêmica e científica proveniente das atividades dos professores e pesquisadores ligados à Universidade. Têm por objetivos, além de reunir a produção acadêmica, preservar e divulgar seus documentos, proporcionando maior visibilidade e aumento do impacto para suas publicações, assim como endossar o Acesso Aberto às produções técnico-científicas. Seu acervo é composto por artigos, relatórios, livros e capítulos de livros, projetos de pesquisa, dissertações, teses e dados de pesquisa. Funciona por meio do protocolo *Open Archives Initiative for Metadata Harvesting* (OAI-PMH) e utiliza a licença *Creative Commons*.

— Portal de dados de Biodiversidade Argentina

O portal é uma iniciativa do Sistema Nacional de Dados Biológicos (SNDB). Oferece um marco de divulgação e acesso aos dados sobre a biodiversidade argentina gerados e/ou administrados pelas instituições aderentes ao Sistema. Seu conteúdo inclui fotos e outros tipos de imagens, material textual, como resenhas e comentários, informações relacionadas à biodiversidade (espécies, morfologia, taxonomia, comportamento, habitat, observações, dados de ocorrência, dados geoespaciais, metadados), bases de dados e conjuntos de dados. Os dados compartilhados por meio do Portal SNDB estão sob as licenças e ferramentas *Creative Commons*. De acordo com as condições que estas estabelecem, qualquer pessoa ou entidade pode fazer uso do Conteúdo.

— Serviço de Difusão da Criação Intelectual (SEDICI) – Repositório Institucional da Universidade Nacional de La Plata

Foi criado em 2003 e, desde então, seu acervo tem crescido exponencialmente, atualmente hospedando milhares de teses de graduação e pós-graduação e dezenas de milhares de artigos e apresentações em conferências dos

membros da UNLP. A SEDICI também acolhe revistas produzidas pelos diversos laboratórios, institutos e centros de investigação que possui a universidade, bem como outras publicações ligadas a ela. O repositório é amparado e regido pelas diretrizes da *Open Archives Initiative* (OAI) e, além de conjuntos de dados, fazem parte do seu conteúdo publicações acadêmicas e científicas, coleções patrimoniais e documentos institucionais.

— Repositório Humadoc

Corresponde ao repositório da Faculdade de Humanidades da Universidade Nacional de Mar del Plata que coleta, preserva e distribui material digital. Por meio do *software* Dspace e do Acesso Aberto, busca alcançar uma maior divulgação de suas pesquisas. Funciona por meio da licença *Creative Commons*.

Brasil:

— *Global Collaboration Engine – Globe*

É um ambiente colaborativo *online* que permite aos pesquisadores de mudança de terras compartilhar, comparar e integrar estudos locais e regionais com dados globais. Utilizam conjuntos de dados globais de variáveis humanas e ambientais por meio de análises de similaridade e representatividade. É um projeto financiado pela *National Science Foundation* conduzido em colaboração com o *Global Land Project* e uma equipe de pesquisadores da Universidade de Maryland e estudantes de Geografia e Sistemas Ambientais (GES), Ciência da Computação e Engenharia Elétrica (CSEE) e Sistemas de Informação (IS). Faz parte da coleção estudos de caso definidos pelos usuários.

— *PPBio Data Repository*

O Programa de Pesquisa em Biodiversidade (PPBio) foi criado em 2004 com o objetivo de aprofundar os estudos sobre biodiversidade no Brasil, descentralizar a produção científica de centros acadêmicos já desenvolvidos, integrar atividades de pesquisa e divulgar resultados para diversos fins, incluindo gestão ambiental e

educação. Os dados se originam de um conjunto altamente distribuído de estações de campo, laboratórios, locais de pesquisa e pesquisadores individuais. O *software* de gerenciamento de dados é desenvolvido de forma gratuita e de código aberto, alimentado pelo sistema de gerenciamento de dados Metacat e otimizado para lidar com conjuntos de dados descritos usando a Linguagem de Metadados Ecológicos, mas pode armazenar qualquer documento de metadados baseado na *Extensible Markup Languageem* (XML). Os dados compartilhados estão sob a licença *Creative Commons*.

— Cariniana Rede Brasileira de Serviços de Preservação Digital – IBICT

O repositório 2Dataverse do IBICT suporta o compartilhamento de dados de pesquisa com citação de dados persistentes, permitindo que eles sejam reproduzidos. Utiliza conjuntos de dados específicos arquivados por colaboradores das instituições participantes da rede Cariniana.

— *Scielo Data*

É um repositório multidisciplinar para depósito, preservação e disseminação de dados de pesquisa⁹ de artigos submetidos ou aprovados para publicação ou já publicados em periódicos da Rede SciELO ou depositados no SciELO *Preprints*. Utiliza como referência os níveis de curadoria usados pelo *CoreTrustSeal* como requisito na avaliação de repositórios de dados confiáveis. Há quatro áreas onde os metadados/arquivos de dados precisam passar por curadoria: arquivos, metadados, termos e versões. Todo tipo de dado que subsidie o artigo submetido ou aprovado para publicação e documentações que facilitem avaliação da pesquisa, reprodução dos resultados e reutilização dos dados de pesquisa pode ser depositado no Scielo Data. Todos os dados são gratuitos para *download* e uso conforme sua Licença

⁹ O Projeto Dataverse é um aplicativo da *web* de código aberto para compartilhar, preservar, citar, explorar e analisar dados de pesquisa. Um repositório Dataverse é a instalação do *software*, que então hospeda vários arquivos virtuais chamados coleções Dataverse. Cada coleção Dataverse contém conjuntos de dados, e cada conjunto de dados contém metadados descritivos e arquivos de dados (incluindo documentação e código que acompanha os dados). O Instituto de Ciências Sociais Quantitativas (IQSS) lidera o desenvolvimento do *software* de código aberto do Projeto Dataverse e, com o *Open Data Assistance Program* em Harvard (uma colaboração com a biblioteca de Harvard, o *Office for Scholarly Communication* e IQSS), fornece suporte ao usuário.

Creative Commons, geralmente CC BY que permite acesso livre, amplo uso e requer citação.

— Centro de Documentação e Acervo Digital da Pesquisa – CEDAP

É um Órgão Auxiliar da Faculdade de Biblioteconomia e Comunicação da Universidade Federal do Rio Grande do Sul. Busca dar suporte para pesquisas científicas, tecnológicas, artísticas e culturais realizada na Universidade, visando à gestão, preservação e uso científico e cultural de seus ativos digitais de pesquisa, natos ou decorrentes da digitalização. Desenvolve políticas aplicadas à gestão do material de pesquisa, seguindo normas internacionais, de forma a propiciar o reuso do material; manutenção dos dados de observação; disponibilização de dados para novos projetos e fomento à colaboração para evitar a duplicação de esforços. Assim o CEDAP, a partir desses recursos, elabora estratégias e planos para digitalização, por meio da seleção, configuração, controle de qualidade e fluxos, elaboração de ontologias, esquemas de metadados e plataformas, concentrando esforços de áreas como a Ciência da Informação, Biblioteconomia, Arquivologia, Comunicação e Ciência da Computação.

— Repositório Institucional UNESP

Busca disseminar, preservar e valorizar suas produções acadêmicas por meio do Acesso Aberto, além de oferecer à sociedade acesso ao conhecimento. Encontra-se organizado em comunidades e subcomunidades que representam as unidades universitárias e auxiliares. Integra o seu acervo artigos, trabalhos e resumos, livros e capítulo de livros, teses e dissertações, notação musical, imagem, material cartográfico, gravação de som e vídeo, *software*, patente, entre outras. Utiliza, preferencialmente, padrões abertos para formatos de arquivos, concedendo acesso livre e gratuito a qualquer usuário. Funciona por meio da licença *Creative Commons*.

— Repositório de dados de pesquisa do Instituto Federal Goiano - Campus Urutaí

Segundo o Instituto Federal Goiano (2020), o projeto “Dados Abertos de Pesquisas” visa disponibilizar dados brutos provenientes de observações e experimentação técnico-científicas, tendo em vista o reuso apropriado de dados em atividades didático-pedagógicas ou em novas pesquisas. Utiliza a plataforma Dataverse e o arquivamento é feito em dataverses, que consistem em *containers* virtuais de arquivos nos quais é possível organizar e disponibilizar dados de pesquisa completamente documentados, seguindo os princípios FAIR. Assim que são depositados, os conjuntos de dados recebem automaticamente e de forma gratuita um identificador digital DOI, permitindo que seja utilizado como referência em publicações científicas.

Chile:

— *Repositorio de datos de investigación de la Universidad de Chile*

Preserva, divulga e dá acesso aos dados de pesquisa gerados por seus acadêmicos e pesquisadores, a fim de dar visibilidade a esses dados, garantir sua preservação e facilitar seu acesso e reutilização. É administrado pela Diretoria de Serviços de Informação e Bibliotecas (SISIB) com base no *software* Dataverse, desenvolvido por Harvard, o qual foi adaptado às necessidades de nossa organização. Os dados da pesquisa podem ser disponibilizados em diversos formatos de arquivos físicos e tipos de conteúdo.

Colômbia:

— *International Center for Tropical Agriculture Dataverse – CIAT*

Conforme afirma o Centro Internacional de Agricultura Tropical, esse é o repositório oficial de produtos e publicações de acesso aberto do CIAT. Em sua coleção, encontram-se artigos de periódicos revisados por pares, resumos de políticas, publicações corporativas, livros, capítulos de livros, manuais, papéis de trabalho, pôsteres, infográficos, vídeos, ferramentas da *web*, apresentações, dentre outros materiais. Os dados de pesquisas produzidos pelo CIAT e seus Parceiros são gratuitos, observando a licença fornecida pelo Dataverse.

— *Repositorio de datos de investigación de la universidad del Rosario*

É uma plataforma institucional que permite o armazenamento, gerenciamento, preservação, disseminação e uso de dados de pesquisa, preparados ou coletados por pesquisadores da Universidad del Rosario. É regido pela política institucional de gestão de dados de pesquisa e utiliza o *software* Dataverse que permite a publicação de dados de pesquisa atribuindo identificadores persistentes (DOI) ajudando a fortalecer a visibilidade das pesquisas científicas.

Espanha:

— *IoChem – BD*

A plataforma de *software* ioChem-BD é um conjunto de serviços *web* orientado para gerenciar todo o ciclo de vida de arquivos digitais de Química Computacional e Ciência de Materiais. Segundo a própria plataforma, conta com um *software* exclusivo, construída de forma modular para satisfazer todas as necessidades: criação e curadoria de dados, publicação, armazenamento, indexação de dados e serviços de mecanismo de busca. O nó principal da rede executa o módulo Find, que atua como servidor central e é alimentado por quaisquer novos dados publicados nos repositórios. O módulo Find fornece um serviço público aberto de mecanismo de busca com conhecimento de produtos químicos rápidos e está hospedado no Barcelona Supercomputer Center (BSC). Outros módulos foram desenvolvidos para automatizar processos de extração de dados relevantes e transformar dados numéricos brutos em dados rotulados em um banco de dados. Também fornece ao pesquisador ferramentas para validar, enriquecer, publicar e compartilhar informações, bem como ferramentas para acessar, pós-processar e visualizar dados.

— *UPCommons*

Como o repositório institucional da Universidade Politécnica da Catalunha, coleta, gerencia, divulga e preserva a produção de ensino e pesquisa dos membros

da comunidade universitária em acesso aberto: artigos de periódicos, relatórios de pesquisa, trabalhos de conferências, teses finais, teses de bacharelado, teses de doutorado, materiais de ensino e outros documentos acadêmicos. É gerido pelo Serviço de Bibliotecas, Publicações e Arquivos e usa principalmente o DSpace, um programa de código aberto desenvolvido pelo *Massachusetts Institute of Technology* (MIT) e pela *Hewlett Packard* (HP). Utiliza-se do protocolo de interoperabilidade *Open Archives Initiative* (OAI-PMH) para aumentar a visibilidade dos documentos depositados, protocolo que permite que outros aplicativos coletem itens de metadados para fazer outros produtos, melhorando sua visibilidade e impacto. Conforme afirma o repositório, os registros de metadados armazenados são protegidos por CC-by 3.0, que autoriza o compartilhamento (cópia e redistribuição em qualquer formato) e a adaptação (remix, transformação e construção sobre o material ou qualquer finalidade, mesmo comercialmente), desde que a autoria seja reconhecida.

— UOC O2 Dades – Universidade Aberta da Catalunha

Criado em 2010 para posicionar a UOC como uma universidade global comprometida com o movimento de acesso aberto. Aceitam arquivos em qualquer formato: aplicativos, vídeo, áudio, etc, porém, a recomendação é que se utilize formatos abertos que garantam a preservação a longo prazo dos materiais, uma vez que são baseados em padrões abertos, livres de restrições legais e econômicas de uso. Cada depositante pode escolher o tipo de licença *Creative Commons* que deseja utilizar.

— Digibug

É o repositório institucional da Universidade de Granada. Recolhe, compila e organiza documentos digitais derivados de suas produções científicas com o intuito de apoiar a investigação científica, ensino e aprendizagem da Universidade. Seu acervo conta com artigos, atas de congresso, conferências, livros, capítulos de livros, teses, trabalhos acadêmicos, recursos educacionais abertos, documentos de trabalho, regulamentos institucionais, patentes e dados de pesquisa.

— RiuNet

Atua como o repositório institucional da Universidade Politécnica de Valência, sendo administrado pela biblioteca. Visa aumentar a visibilidade sobre sua produção acadêmica e científica, assim como torna-la acessível e preservá-la. Utiliza a licença *Creative Commons* para distribuir e conceder acesso a seu acervo que conta com objetos de aprendizagem, teses, artigos, mapas, trabalhos acadêmicos, patrimônio institucional, multimídia, material didático, revistas e anais de congresso divididos em quatro áreas correspondentes à Docência, Investigação, Institucional e Patrimonial. Os registros dos materiais contam com metadados que informam sobre a possibilidade de acesso ao seu conteúdo. Também utiliza a licença *Creative Commons* para compartilhamento de seus materiais. Recomenda aos seus depositantes que utilizem o inglês, espanhol e valenciano para seus descritores e resumos com a finalidade de favorecer a recuperação dos documentos.

— Repositório Digital de Documentos UAB

Em 2006, o Serviço de Bibliotecas lançou o Repositório Digital de Documentos UAB com o objetivo de coletar, organizar e preservar toda a documentação digital que chega às bibliotecas seja científica, patrimonial ou institucional. O Repositório Digital de Documentos da Universidade Autônoma de Barcelona é a ferramenta a partir da qual é coletada, gerenciada, divulgada e preservada a produção científica, docente e institucional da universidade, ao mesmo tempo em que coleta documentos digitais que fazem parte dos acervos. O Serviço de Biblioteca da UAB é responsável pela divulgação, promoção e captação de conteúdo para o repositório e o Serviço de Informática da UAB é responsável pelo funcionamento do repositório e pela preservação digital dos seus conteúdos. O acesso aos registros e metadados é gratuito e aberto. O texto integral dos conteúdos pode ser de acesso aberto ou acesso restrito ao campus da UAB. Utiliza a licença *Creative Commons*.

México:

— Centro Internacional de Melhoramento de Milho e Trigo (CIMMYT) – Rede

de repositório de dados e *software* de pesquisa

O CIMMYT cresceu a partir de um programa piloto patrocinado pelo governo mexicano e pela Fundação Rockefeller nas décadas de 1940 e 50, com o objetivo de aumentar a produtividade agrícola no México. Disponibiliza um repositório de acesso aberto e gratuito de *softwares* de pesquisa por meio do Dataverse, estudos e conjuntos de dados produzidos e desenvolvidos por cientistas do CIMMYT, bem como os resultados do projeto *Seeds of Discovery*, que disponibiliza perfis genéticos de trigo e milho.

— cIRcle

É o repositório digital de acesso aberto da Universidade de Colúmbia Britânica que reúne materiais da instituição e de parceiros. Busca compartilhar e preservar a produção intelectual exclusiva da Universidade, tornando o conteúdo disponível gratuitamente para qualquer pessoa, em qualquer lugar através da web. O cIRcle faz parte de uma iniciativa para desenvolver um sistema que tornasse as pesquisas acessíveis em repositórios de acesso aberto. Utiliza uma licença de distribuição não exclusiva (v. 3.0), por meio da qual tem o direito de copiar, distribuir e transmitir os materiais pela Internet.

— RU-Econômicas - RU-IIEc

É o repositório universitário do Instituto de Investigação Económica (IIEc) da Universidade Nacional Autónoma do México (UNAM), criado para gerir, divulgar e preservar, em formato digital, a produção intelectual do Instituto de Investigação Económica. Tem por objetivo incentivar a comunicação académica e aumentar a visibilidade e reuso dos materiais produzidos pelo Instituto. Seu acervo é formado por livros, revistas, artigos, conferências, apresentações, audiovisuais, dados de pesquisa, entre outros. O repositório faz parte de um projeto intitulado *UNAM Digital Collections Network (RAD)* que visa criar uma rede de repositórios universitários para apoiar as dependências universitárias na gestão e divulgação dos seus recursos digitais. Utiliza o *software* livre *EPrints* desenvolvido na Escola de Eletrônica e Ciência da Computação da Universidade de Southampton na Inglaterra.

Portugal:

— *DataRepositorium* - Repositório de dados da universidade do Minho

Constituído no ano de 2020 com o objetivo de armazenar, preservar, divulgar e dar acesso à produção intelectual da Universidade do Minho em formato digital, o repositório reúne num único sítio o conjunto das publicações científicas académicas da comunidade pertencente à instituição. Utiliza-se do Dataverse para compor seu repositório.

— *Kinetic models of biological systems – KiMoSys*

É uma plataforma de acesso livre na área de biologia de sistemas baseada na web. Possui uma aplicação *web* para modelos quantitativos de sistemas biológicos. Combina tarefas para armazenamento de dados experimentais, pesquisa e compartilhamento e ferramentas para construir modelos cinéticos baseados em ODE em um único lugar e se destina a apoiar a comunidade de biologia de sistemas ao fazer pesquisas experimentais e computacionais. Suporta variados formatos padrão de dados. Armazena dados de séries temporais de metabólitos, metabólitos em estado de equilíbrio, dados de fluxo e níveis de proteína. O repositório é protegido por direitos autorais da equipe que desenvolve o recurso sob licença GNU GPL (v2)¹⁰. No entanto, cada arquivo de dados e modelo individual mantém os direitos autorais atribuídos pelos biólogos que mensuram os dados e pelos criadores do modelo cinético, bem como pelo autor da publicação. Além disso, quando os arquivos de dados (e/ou modelo) são compartilhados pelo remetente como públicos, eles entendem que os dados públicos (e/ou modelo) estão disponíveis gratuitamente por qualquer outro usuário e para qualquer finalidade sob a licença *Creative Commons*.

— DadosIPB

¹⁰ As licenças GPL são utilizadas por projetos de software livre e de código aberto. Disponível em: <http://www.gnu.org/licenses/old-licenses/gpl-2.0.html>.

É o repositório do Instituto Politécnico de Bragança. Tem como objetivo armazenar, preservar, compartilhar e gerenciar dados de pesquisa gerados por sua comunidade acadêmica. Utiliza o *software Dataverse* e funciona por meio da licença CC0 1.0 – Universal (CC0 1.0) Dedicção ao Domínio Público. Possui o acervo composto por livros, artigos, *papers* de conferência, teses, dissertações, relatórios, dados de pesquisa, entre outros.

A partir da exposição pormenorizada dos procedimentos metodológicos e do universo de pesquisa, recorre-se às discussões provenientes da literatura especializada da Biblioteconomia e Ciência da Informação para estabelecer a fundamentação teórica desta pesquisa, apresentada na próxima seção.

3 REFERENCIAL TEÓRICO

Nesta seção, são contextualizados, à luz dos referenciais teóricos provenientes da biblioteconomia e Ciência da Informação, os temas abordados nesta pesquisa. Inicialmente, trabalha-se com o movimento de acesso aberto e as principais iniciativas e estratégias que sustentam essa prática e filosofia na atualidade. Idealmente, inserem-se discussões sobre repositórios digitais, com destaque para o repositório de dados de pesquisa, dados de pesquisa e a prática do autoarquivamento. Na sequência, integram-se reflexões que versam sobre o metadado assunto em tempos de autoarquivamento, tendo como questão central o papel assumido pelo autor-indexador na atribuição de palavras-chave com fins de representação dos conteúdos dos recursos informacionais, e os benefícios e implicações dessa prática na representação da informação para fins de recuperação por assunto.

3.1 O MOVIMENTO DE ACESSO ABERTO

Com o desenvolvimento constante das tecnologias de informação e comunicação, a produção e disponibilização de dados e documentos em formato digital se expandiram, reivindicando novas formas pelas quais a comunicação científica seria exercida. Em vista disso, surge o movimento de Acesso Aberto que desafia antigos paradigmas da ciência e se torna um tema relevante para discussões que envolvem a produção técnico-científica.

Costa (2006) define o movimento de Acesso Aberto como um movimento voltado para o uso de ferramentas, estratégias e metodologias que indicam uma nova forma de representar essa nova modalidade de disponibilização da comunicação científica. Abrange, entre outras coisas, *software* aberto (livre), arquivos abertos (a fim de se alcançar a interoperabilidade em nível global) e o próprio Acesso Aberto que contribui com a disseminação irrestrita de resultados de pesquisas científicas. Nesta esfera, a Ciência Aberta se constitui em

Um movimento que incentiva a transparência da pesquisa científica desde a concepção da investigação até o uso de softwares abertos. Também promove esclarecimento na elaboração de metodologias e gestão de dados científicos, para que estes possam ser distribuídos, reutilizados e estar acessíveis a todos os níveis da sociedade, sem

custos. Propõe, ainda, a colaboração de não cientistas na pesquisa, ampliando a participação social por meio de um conjunto de elementos que dispõem de novos recursos para a formalização da comunicação científica (SILVA; SILVEIRA, 2019, p. 2).

Foi em meio a essas ações que se desenhou o movimento do Acesso Aberto, o qual apoia que toda e qualquer informação científica se encontre disponível de forma integral para ser acessada a quem interessar. Nesse âmbito, as principais iniciativas que corroboram com a discussão da temática, destacando-se sobretudo o uso de *softwares* e a disponibilização de arquivos abertos, são a *Budapest Open Access Initiative* (2002), reunião de Bethesda (2003) e a Declaração de Berlim (2003).

A *Budapest Open Access Initiative* (BOAI) foi uma das primeiras iniciativas a estabelecer o termo “*open access*” com o objetivo de incentivar o avanço da Ciência Aberta. Determina duas estratégias principais fundamentadas no uso do *Open Archives Initiative – Protocol for Metadata Harvesting* (OAI-PMH) para se alcançar esse objetivo. A primeira consiste na prática do autoarquivamento que, segundo a Declaração de Budapeste,

Os acadêmicos precisam das ferramentas e da assistência para depositar seus artigos de periódicos avaliados em arquivos eletrônicos abertos [...] quando esses arquivos estão em conformidade com os padrões criados pela *Open Archives Initiative*, os mecanismos de pesquisa e outras ferramentas podem tratar os arquivos separados como um só. Assim, os usuários não precisam saber quais arquivos existem ou onde estão localizados para encontrar e fazer uso de seus conteúdos (BUDAPEST OPEN ACCESS INITIATIVE, 2002, n. p., tradução nossa).

Essa primeira estratégia é definida como Via Verde ocorrendo quando “tais artigos, geralmente na forma de manuscritos do autor que precederam o artigo finalizado, são disponibilizados gratuitamente em algum lugar na web.” (BJÖRK et al., 2014, p. 3, tradução nossa). Os autores autoarquivam seus trabalhos que já foram publicados em um periódico científico. Logo, os editores que aceitaram esses trabalhos concedem permissão para o depósito dos artigos em repositórios digitais também de acesso aberto. Essa prática traz uma série de benefícios não somente para os usuários dos ambientes digitais, mas também para os autores das pesquisas, pois tornam acessíveis as publicações de artigos científicos que somente

estariam disponíveis em periódicos científicos que necessitam de assinatura.

A segunda estratégia refere-se aos periódicos de acesso aberto, estabelecendo a Via Dourada. Para Veiga e Macena (2015) a Via Dourada se subdivide de duas formas, podendo ser pura, constituindo-se de periódicos publicados por instituições e sendo escolha do próprio pesquisador publicar seus trabalhos em periódicos de acesso aberto, onde os artigos são disponibilizados *online*. Porém, pode se estabelecer também por meio de periódicos publicados por editores comerciais, da mesma forma, sendo do pesquisador a escolha de publicar em periódicos científicos que disponibilizam os artigos *online*, contudo há o pagamento de taxas pelo próprio autor ou instituição. Já na Via Dourada híbrida tem-se um periódico de acesso aberto e acesso restrito. O pesquisador pode escolher publicar seus estudos em periódicos científicos que disponibilizam somente alguns artigos em acesso aberto, por meio de pagamento de taxas pelo pesquisador ou pela instituição.

Essas práticas se consolidam por meio do que estabelece a BOAI quando elucida que

Os acadêmicos precisam de meios para lançar uma nova geração de periódicos comprometidos com o acesso aberto e para ajudar os periódicos existentes que optam por fazer a transição para o acesso aberto. Como os artigos de periódicos devem ser disseminados o mais amplamente possível, esses novos periódicos não mais invocarão os direitos autorais para restringir o acesso e o uso do material que publicam. Em vez disso, eles usarão direitos autorais e outras ferramentas para garantir acesso aberto permanente a todos os artigos que publicam. (BUDAPEST OPEN ACCESS INITIATIVE, 2002, n. p., tradução nossa).

A reunião de Bethesda legitima o que foi instituído pela BOAI quando discute como alcançar o objetivo de levar o acesso aberto à produção científica. Para esse fim, foi estabelecido nessa reunião o que são publicações de acesso aberto, que abrange dois pontos principais. O primeiro ponto se refere a direitos autorais, onde os autores devem garantir aos usuários direito livre, irrevogável e perpétuo de acesso aos seus trabalhos, como também

Licença para copiar, usar, distribuir, transmitir e exibir trabalhos publicamente, elaborar e distribuir trabalhos derivados, em qualquer meio digital, para qualquer propósito responsável, sujeito à atribuição apropriada de autoria, assim como fazer um número pequeno de

cópias impressas para uso pessoal. (COSTA, 2006, p. 42).

O segundo ponto elucida que deve ser feito o depósito integral do trabalho, incluindo qualquer material que foi utilizado para os fins da pesquisa, como também uma cópia da licença para uso e distribuição. Esse depósito deve ser feito em um repositório que disponibilize o material de forma *online*, que seja vinculado a uma instituição acadêmica, científica ou governamental que conceda acesso livre e garanta a interoperabilidade a fim de que o arquivamento do material ocorra em longo prazo, consecutivamente à publicação do trabalho original.

Quanto à Declaração de Berlim, a reunião de 2003 que discutiu o acesso aberto às ciências e humanidades trouxe a Internet como ator principal que opera como o pilar do conhecimento científico global. Da mesma forma, discutiu a respeito das diretrizes a serem consideradas pelas instituições a fim de que se atingisse o acesso aberto, incluindo os museus como uma representante das instituições de patrimônio cultural. Esse fato veio da decorrência de ter-se debatido não somente o conhecimento gerado pela pesquisa científica, mas também sobre a herança cultural.

Durante a reunião de 2005 foi posto como contexto para a implementação da Declaração de Berlim que as instituições acompanhassem os seguintes pontos:

1. Implementar uma política para exigir que seus pesquisadores depositem uma cópia de todos os seus artigos publicados em um repositório de acesso aberto.
2. Incentive seus pesquisadores a publicar seus artigos de pesquisa em revistas de acesso aberto onde exista uma revista adequada e forneça o suporte para que isso aconteça (SOCIEDADE MAX PLANCK, 2005, n. p., tradução nossa).

Essas foram as principais iniciativas e estratégias que impulsionaram as instituições e agências de fomento a estabelecerem como parte de suas políticas que os pesquisadores depositassem suas pesquisas em ambientes digitais para serem disponibilizadas *online*, assim como elaborassem plano de gestão de dados, corroborando com o movimento em prol do acesso aberto. Entretanto, tem-se em destaque especial o autoarquivamento que ainda hoje, mesmo com o crescimento exponencial dos repositórios digitais, encontra barreiras para a sua implementação e utilização em ambientes digitais.

O autoarquivamento foi apresentado para a comunidade científica pela

primeiravez em 1995 por Stevan Harnad, mediante uma postagem *online* chamada de “Proposta Subversiva”. Para Veiga e Macena (2015) esta é uma prática que permite ao pesquisador por si só tornar sua pesquisa acessível para qualquer usuário que possua acesso à internet, de forma que a visibilidade da sua pesquisa seja ampliada. Contudo, o autoarquivamento se mostra como uma prática que não acontece de forma instintiva, sendo necessário que as instituições e agências de fomento estabeleçam políticas que incentivem o auto-depósito.

A fim de debater sobre os propósitos do autoarquivamento, Swan (2008) afirma não se tratar de uma alternativa à publicação em periódicos científicos, mas sim uma atividade complementar em que um autor publica seu artigo em qualquer periódico que escolher e, em seguida, arquiva uma cópia por conta própria. Os arquivos auto-depositados são classificados como *postprints* revisados por pares ou rascunhos de *preprint*, o que ratifica a definição de via verde evidenciada por Björk et. al. (2014), para quem só há efetivação da via verde quando os estudos são avaliados por pares.

A criação de repositórios de dados deve ser acompanhada de diretrizes explícitas, pois quando essas diretrizes não se encontram bem estabelecidas surgem obstáculos no compartilhamento das pesquisas, principalmente no que diz respeito às licenças.

Para Medeiros (2010), apesar da agilidade na disponibilização da informação, as tecnologias digitais não resolvem todos os problemas, sendo muitas as implicações do autoarquivamento na qualidade da representação e recuperação por assunto em sistemas abertos de informação científica. Conforme evidenciado, a pesquisa de Freitas, Dal'Evedove e Tartarotti (2021) confirmou esse cenário ao constatar que dos mais de cem repositórios institucionais brasileiros listados no IBICT, apenas um traz em sua política de autoarquivamento indicações sobre a qualidade dos metadados, porém sem menção ao metadado assunto.

Para Veiga e Macena (2015), a baixa adesão ao autoarquivamento entre pesquisadores e instituições se deve a multifatores que precisam ser avaliados conforme cada instituição e/ou área do conhecimento. Para isso, devem ser consideradas como medidas a redução de barreiras e criação de incentivos para que esta prática possa ter sucesso. Para tanto, a discussão da temática com a comunidade científica se mostra um possível caminho para a supressão da resistência à prática por parte do autor-pesquisador, de modo que haja maior adesão

quanto ao compartilhamento das pesquisas de forma integral, como também dos dados de pesquisa que são gerados por esses estudos.

3.2 REPOSITÓRIOS DE DADOS DE PESQUISA

De maneira a colaborar com o movimento de acesso aberto, os repositórios digitais atuam como bases de dados online que reúnem a produção técnico-científica de uma instituição ou área temática, conforme os seus propósitos. Diante da necessidade de armazenar, preservar, organizar e compartilhar os documentos para que se tornem acessíveis a todo e qualquer usuário, as universidades, bibliotecas, arquivos, museus e agências de fomento, entre outras instituições interessadas em progredir por meio da ciência, viram os repositórios digitais como uma oportunidade para se desenvolverem, ao mesmo tempo como forma de apoiarem o movimento de acesso aberto e aperfeiçoar os serviços fornecidos. Os repositórios digitais se tornam peças centrais no contexto da Ciência Aberta ao reunirem de forma organizada informações resultantes do meio científico.

Com o intuito de abarcar o maior número de áreas possível, várias tipologias de repositórios digitais surgiram, visto que cada comunidade opta por utilizar seu próprio modelo de comunicação científica. É desse meio que emergem os repositórios institucionais, cujos resultados positivos desencadearam outras tipologias de repositórios, em virtude de “serem orientados para disseminar a produção intelectual criada no âmbito das instituições; sendo planejados, desenvolvidos, implantados e mantidos por elas” (VILLALOBOS; GOMES, 2015, p. 9).

Os repositórios institucionais surgem como uma maneira que as instituições e universidades encontraram para aumentar a visibilidade sobre sua produção técnico-científica. Cada vez mais países ao redor do mundo têm aderido a essa ferramenta, que não só proporciona credibilidade às informações produzidas, como também favorece o processo de produção de conhecimento que busca se aprimorar em razão das possibilidades advindas com as Tecnologias de Informação e Comunicação.

Lynch (2003) evidencia que em 2003 a Fundação *Andrew W. Mellon*, juntamente com outras instituições, financiou o DSpace, que foi desenvolvido pelo Instituto de Tecnologia de *Massachusetts* (MIT) e programado para ser

implementado em várias instituições no mundo. Por ser um *software* de código aberto, contribuiu para que os custos e entraves para desenvolvimento e implementação de repositórios institucionais fosse reduzido significativamente. Para o referido autor, os repositórios institucionais são responsáveis por desempenhar várias funções e tem um papel central na transformação do conhecimento científico.

Um repositório institucional universitário é um conjunto de serviços que uma universidade oferece aos membros de sua comunidade para a gestão e divulgação de materiais digitais criados pela instituição e seus membros da comunidade. É essencialmente um compromisso organizacional com a administração desses materiais digitais, incluindo preservação a longo prazo, quando apropriado, bem como organização e acesso ou distribuição (LYNCH, 2003, p. 328, tradução nossa).

Nesse sentido, os repositórios institucionais representam mais do que somente *softwares* e *hardwares*, mas também a garantia de que o conhecimento científico seria preservado, pois os repositórios institucionais se tornavam responsáveis pela gestão das mudanças tecnológicas e a migração dos conteúdos digitais para o próximo conjunto de tecnologias que surgisse, sendo parte das atribuições organizacionais que fornece.

A criação de repositórios institucionais pelas universidades e organizações alavancou o progresso de outras tipologias de repositórios, sendo eles fundamentais para o desenvolvimento e aperfeiçoamento dos instrumentos e serviços oferecidos por esses ambientes digitais, ao mesmo tempo em que eram observadas algumas dificuldades. Como exemplo, os dados de pesquisa, que cada vez mais atuam como importantes recursos de divulgação científica, não encontraram nesses ambientes formas de serem englobados. Assim, os repositórios de dados de pesquisa emergem com a função de lidar com todas as ações que envolvem os dados de pesquisa, sua preservação e compartilhamento.

Os repositórios de dados de pesquisa auxiliam no gerenciamento, armazenamento, preservação, disseminação e acesso de dados que “são evidências e/ou insumos de pesquisa [...] coletados, observados, registrados e/ou criados para fins de análise e que podem produzir resultados de pesquisa para um estudo científico” (SEMELER; PINTO, 2019, p. 119). Constituem sistemas de informação que podem reunir arquivos de diversos formatos e trazer benefícios tanto ao pesquisador, como também às instituições e comunidades científicas, tendo

como principal produto os dados de pesquisa. Em vista disso, os repositórios de dados de pesquisa surgem como

Infraestruturas de base de dados desenvolvidas para apoiar todo o ciclo da gestão de dados de pesquisa, incluindo as ações mais dinâmicas e contundentes sobre os dados, que coletivamente são chamadas de curadoria de dados de pesquisa e visam adicionar valor aos dados, avaliando, formatando, agregando e derivando novos dados (SAYÃO; SALES, 2016).

Para Gama et al., (2017), os repositórios de dados de pesquisas têm a finalidade de mediação informacional ao aproximar instituições e autores-pesquisadores por meio do compartilhamento da informação científica. Buscam adotar sistemas para o desenvolvimento e administração de dados digitais, associados às políticas que asseguram qualidade, acesso e compartilhamento, para que se tenha maior visibilidade sobre os resultados das pesquisas e possibilite a preservação da memória científica.

O ciclo de vida dos dados é regido pela gestão dos dados de pesquisa. São ações que, segundo Sayão e Sales (2016), se iniciam no planejamento da criação dos dados, passam pela organização em coleções identificadas por meios de referências estáveis e padronizadas e vão até o arquivamento de longo prazo para os dados reconhecidos como de valor permanente ou que são referências universais.

Os benefícios dos repositórios de dados de pesquisa se constituem em:

- 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. 63).

Diante da constante geração de dados de pesquisa, observa-se o desenvolvimento dessas infraestruturas que são capazes de garantir o acesso e, conseqüentemente, reuso desse tipo de recurso informacional em ambientes tecnológicos que são gerenciáveis e confiáveis. Neste âmbito, os repositórios de dados de pesquisa têm feito papel de mediadores entre pesquisadores e instituições, fornecendo segurança e qualidade, sendo ponte para o acesso às pesquisas e garantido o compartilhamento científico.

Rodrigues et. al. (2010) elucida que os repositórios de dados de pesquisa passam por três níveis de curadoria, designados como armazenamento, representação do conjunto e representação do item:

O primeiro nível, do armazenamento, trata de garantir que o conjunto de dados está guardado de forma fiável, num sistema com garantia de manutenção e salvaguarda. O nível 2, da representação do conjunto, assegura que existe descrição do conjunto de dados, garantindo metadados tais como os que dizem respeito à sua produção, direitos de uso ou características técnicas. O nível 3, da representação do item, lida com as questões de modelo e descrição para itens individuais e é dependente da natureza do conjunto de dados (RODRIGUES et. al., 2010, p. 23).

São observados dois tipos de infraestruturas para suporte desses ambientes, em que a primeira permeia repositórios de dados de pesquisa que já consolidaram práticas de registro e compartilhamento de dados. Neste caso, ocorrem porque são estabelecidos padrões para a publicação ou porque não há meios para coletar dados próprios e a investigação fica restrita aos dados recolhidos por órgãos supervisores. Conforme declara o autor, “o paradigma vigente é a base de dados, e não o repositório de dados. Isto significa que, dos três níveis de curadoria, estão assegurados o 1 (armazenamento) e o 3 (representação do item)” (RODRIGUES et. al., 2010, p. 24).

O segundo tipo de infraestrutura para suporte relaciona-se ao movimento de acesso aberto, em que as instituições são compelidas a disponibilizarem seus dados coletados ou projetos que foram financiados por agências de fomento ou obtiveram recursos públicos. Aqui, os conjuntos de dados se encontram associados às publicações, sendo o elemento principal o uso das plataformas dos repositórios. As mais conhecidas

São as de desenvolvimento com licenças de código aberto, que surgiram para dar resposta à necessidade de colecionar e preservar a produção de natureza científica e técnica dentro de comunidades de investigação. O seu uso para os repositórios de dados dá ênfase aos níveis de curadoria 1 e 2, uma vez que são capazes de lidar com o armazenamento e a descrição ao nível da coleção, mas não com a descrição ao nível do item (RODRIGUES et. al., 2010, p. 24).

Nos repositórios de dados de pesquisa são aplicadas várias soluções tecnológicas que ora são desenvolvidas de forma geral para esses ambientes, ora são desenvolvidas a partir de casos específicos que dependerão da categoria dos conjuntos de dados e do uso que será feito a partir deles. Dessa forma, busca-se alcançar os objetivos da Ciência Aberta por meio de ações técnicas que partem desse movimento.

A interoperabilidade é uma das ações técnicas que consiste na “comunicação entre diferentes sistemas de informação, condição na qual os sistemas necessitam disponibilizar os metadados, os quais expõem informações fundamentais” (VILLALOBOS; GOMES, 2015, p. 4), garantindo a identificação, descrição e localização de recursos informacionais que se encontram dispostos em ambientes digitais, como no caso dos repositórios de dados de pesquisa. No entanto, como ação isolada a favor da encontrabilidade da informação, se mostra insuficiente, sendo fundamental uma padronização da representação da informação em ambientes digitais, sejam ou não colaborativos.

3.3 DADOS DE PESQUISA

A palavra “dados” se origina do latim “datum” que tem por significado “aquilo que foi dado”, podendo ser definido ainda como quaisquer registros armazenados em um computador (BUCKLAND, 1991). Para Johnston (2017), dados podem ser qualquer informação factual e que possa ser analisada. Contudo, o termo pode adotar vários sentidos de acordo com o contexto em que está sendo utilizado.

Em um sentido mais amplo, o *National Research Council* (1999) afirma que dados podem ser definidos como “fatos, números, letras e símbolos que descrevem um objeto, ideia, condição, situação ou outros fatores”. Por sua vez, a *Organization for Economic Co-Operation and Development* (2007) salienta que dados são “registros de fatos usados como fontes primárias na investigação científica e que

geralmente são aceitos na comunidade científica como necessários para a validação dos resultados da pesquisa.” Para Davis e Vickery (2007), os dados se constituem na atualidade como o recurso mais valioso da ciência.

Para os fins deste trabalho, adota-se o entendimento de que “dado de pesquisa é todo e qualquer tipo de registro coletado, observado, gerado ou usado pela pesquisa científica, tratado e aceito como necessário para validar os resultados da pesquisa pela comunidade científica” (SALES; SAYÃO, 2019, p. 36). Considera-se que esta definição engloba todos os fatores a serem observados quando se reflete sobre o conceito de dados, sendo assim a mais adequada no contexto desta pesquisa.

No estudo intitulado ‘Uma proposta de Taxonomia para Dados de Pesquisa’, Sales e Sayão (2019, p. 7, grifo nosso) expandem as tipologias de dados de pesquisa os classificando a partir de diversos aspectos, quais sejam:

- **Quanto à natureza:** Número. Medida. Resultado de levantamento. Resultado de experimento. Fórmula. Equação. Algoritmo. Software. Base de dados. Simulação. Códigos nucleares. Multimídia. Imagem. Vídeo. Áudio. Animação. Filme. Fotografia. Visualização. Tabelas. Gráficos. Diagramas. Modelo. Modelo de representação. Metadado. Modelo em 3D. Modelo reduzido. Desenho. Textuais. Questionário. Entrevista. Anotação. Norma. Padrão. Certificado. Caderno de laboratório. Transcrição. Correspondência. Diário. Caderno de campo. Artefato. Espécimes. Amostras. Maquetes. Processo. Procedimento operacional padronizado. Workflow. Protocolo. Teste;
- **Quanto ao grau de processamento ou estágio de geração:** Primário, cru ou brutos. Intermediários, derivados ou pré-processados. Finais ou processados. Terciários, condensados ou de alta densidade;
- **Quanto à origem:** Observacionais. Experimentais. Computacionais. Registro. Governamentais. Dados oriundos de redes sociais;
- **Quanto à formação das coleções:** Coleção de pesquisa. Coleção de comunidade. Coleção de referência;
- **Quanto ao nível de abertura:** Fechado. Embargado. Parcialmente aberto. Aberto;
- **Quanto ao nível de sensibilidade:** Dados pessoais. Dados pessoais sensíveis. Dados confidenciais;
- **Quanto à mutabilidade:** Canônico. Episódico;
- **Quanto à materialidade:** Físico. Digital.

Em contrapartida, em estudo mais recente, os mesmos autores classificam dados de pesquisa em observacionais, computacionais e experimentais, nos quais:

- **Observacionais:** provém da observação de fenômenos e acontecimentos únicos. Esses dados “são tipicamente coletados por meio de percepção humana ou mensuração (por exemplo, notas de campo) ou, mais comumente, por sensores ou por outros instrumentos” (SAYÃO; SALES, 2020). Podem ser coletados de acordo com duas categorias, sendo *in situ* (quando o pesquisador tem contato direto com o fenômeno) ou em contraste (quando é realizado à distância através de instrumentos);
- **Computacionais:** são dados resultantes de processamentos computacionais, simulações e *workflow*. Esse tipo de dados “não precisam de preservação de longo prazo em repositório ou centros de dados, posto que os programas podem ser executados novamente e os dados reproduzidos” (SAYÃO; SALES, 2020); e
- **Experimentais:** são dados originados a partir de procedimentos controlados a fim de se verificar hipóteses. Contudo, “nem sempre é possível reproduzir precisamente todas as condições experimentais, especialmente quando algumas condições e variáveis experimentais podem não ser conhecidas” (SAYÃO; SALES, 2020).

A partir da chegada da Internet e do desenvolvimento das Tecnologias de Informação e Comunicação (TICs) houve uma multiplicação no volume de dados, o que levou a novos meios pelos quais os dados podem ser acessados, coletados, organizados, armazenados e compartilhados, modificando também a forma como a ciência e os pesquisadores se comunicam e compartilham seus dados.

Nesta esfera, Swanson e Rinehart (2016) elencam vários serviços oferecidos por bibliotecas que se originaram dessa modernização das TICs e que também podem ser estendidos a outras instituições ou espaços, como os repositórios de dados de pesquisa. Os serviços mencionados incluem o planejamento para gerenciamento de dados, assistência com coleta de dados de fontes de texto, aconselhamento sobre documentação de dados e metadados, demonstração do impacto da publicação do conjunto de dados e fornecimento de ferramentas para descoberta, acesso e preservação do conjunto de dados.

Os dados necessitam ser avaliados para que se definam questões como armazenar ou não, tempo de armazenamento, de que forma serão tratados e a que

público serão direcionados. Dificilmente os dados são utilizados em sua forma bruta, ocorrendo a curadoria que origina dados derivados que passaram por vários processos de refinamento e análise de acordo com os objetivos da pesquisa e do pesquisador. Ainda assim, os dados brutos são fundamentais para a reprodutibilidade de pesquisas científicas, uma vez que não passaram por nenhum nível de processamento e contém informações integrais.

Desta forma, os dados de pesquisa tornam-se indispensáveis para o avanço da ciência, colaborando com a validação de estudos e a condução de novas variações de pesquisas, fazendo-se necessário o gerenciamento correto desses ativos para a disponibilização em repositórios de dados de pesquisa, de forma que o compartilhamento com outros pesquisadores seja eficaz.

3.4 OS PRINCÍPIOS *FAIR* EM REPOSITÓRIOS DE DADOS DE PESQUISA

Os princípios *FAIR*, cujo acrônimo significa *Findable* (Localizáveis), *Accessible* (Acessíveis), *Interoperable* (Interoperáveis) e *Reusable* (Reutilizáveis), começaram a ser discutidos em 2014, quando a instituição holandesa *Elixir* juntamente com o *Dutch Techcentre for Life Sciences* (DTL), *Netherlands eScience Center* e o *Lorentz Center* realizaram a conferência “*Jointly designing a Data FAIRPORT*”. Nos quatro dias que se seguiram, editores, especialistas, cientistas da computação e instituições de pesquisa discutiram sobre os principais aspectos de uma infraestrutura global para publicação, descoberta, compartilhamento e reutilização de dados que seria explorada no contexto das *eScience*.

As discussões envolveram aspectos a serem considerados para o desenvolvimento dessa infraestrutura, tais como o contexto, as necessidades do usuário, os desafios tecnológicos, o acesso e os modelos de negócios, os requisitos de financiamento e a governança dessa iniciativa global. Como resultado dos debates, os participantes concordaram que uma infraestrutura global para publicação, descoberta, troca e reutilização de dados é indispensável para uma pesquisa científica eficaz que se baseia em dados, assim como para atender as requisições exigidas por agências financiadoras.

O principal resultado a ser considerado foi a interoperabilidade global de dados de pesquisa. Segundo a *Dutch Techcentre for Life Sciences* (2014), os computadores precisam identificar de maneira independente todos os conjuntos de

dados que se encontrarem disponíveis para um requisito de pesquisa específico. Desse modo, a peça central dessa infraestrutura é um protocolo mínimo que define a interoperabilidade semântica básica de conjuntos de dados de pesquisa por meio de padrões e protocolos comunitários já estabelecidos ou despontantes.

Essa conferência buscou apoiar as comunidades científicas existentes possibilitando um ambiente em que os dados oriundos de pesquisas científicas sejam justos no sentido de serem localizáveis, acessíveis, interoperáveis e reutilizáveis. Esses princípios foram projetados para serem aplicados por protocolos padrão minimamente necessários com interfaces leves e formatos definidos. O objetivo final se constitui em oferecer “modelos mínimos para agrupar resultados e vincular dados a análises que sejam acionáveis por humanos e por computador” (DATA FAIRPORT, 2014).

Os princípios *FAIR* se preocupam em fornecer um conjunto de metadados que seja utilizado por computador e por pessoas. Quando são implementados de forma correta proporcionam a interoperabilidade entre diferentes sistemas que utilizam dados. Wilkinson et al. (2016) atribui aos princípios *FAIR* um conjunto de funções que são descritas na Tabela 3, quais sejam:

Tabela 3 – Os princípios FAIR.

Para ser localizável (Findable)	F1. Metadados são atribuídos a um identificador globalmente exclusivo e persistente.
	F2. Os dados são descritos com metadados avançados.
	F3. Metadados incluem clara e explicitamente o identificador dos dados que descreve.
	F4. Metadados são registrados ou indexados em um recurso pesquisável.
Para ser acessível (Accessible)	A1. Metadados são recuperáveis por seu identificador usando um protocolo de comunicação padronizado.
	A1.1. O protocolo é aberto, gratuito e universalmente implementável.
	A1.2. O protocolo permite um procedimento de autenticação e autorização, quando necessário.
A2. Metadados são acessíveis, mesmo quando os dados não estão mais disponíveis.	
Para ser interoperável (Interoperable)	I1. Metadados usam uma linguagem formal, acessível, compartilhada e amplamente aplicável para a representação do conhecimento.

	I2. Metadados usam vocabulários que seguem os princípios <i>FAIR</i> .
	I3. Metadados incluem referências qualificadas a outros metadados.
Para ser reutilizável (Reusable)	R1. Metadados são ricamente descritos com uma pluralidade de atributos precisos e relevantes.
	R1.1. Metadados são liberados com uma licença de uso de dados clara e acessível.
	R1.2. Metadados estão associados à proveniência detalhada.
	R1.3. Metadados atendem aos padrões da comunidade relevantes para o domínio.

Fonte: (WILKINSON et.al, 2016; traduzido e adaptado pela autora)

De maneira a caracterizar esses princípios *FAIR* de forma mais específica, apresenta-se a seguir descrições de cada função indicada no Quadro 1, a saber:

a) Localizáveis (*Findable*) – pessoas e computadores precisam encontrar os dados. Desse modo, os metadados precisam ser reconhecíveis para essas máquinas, sendo fundamentais para os conjuntos de dados serem descobertos.

- **F1:** os metadados precisam ter um único identificador (PID, *persistent identifier*), como o DOI. É considerado o atributo mais essencial, posto que “sem identificadores globais únicos e persistentes será difícil conseguir outros elementos de dados *FAIR*” (HENNING et al., 2019, p. 400);
- **F2:** os metadados precisam ser enriquecidos o suficiente para que seja possível para o sistema realizar várias ações automaticamente. Dessa forma, o usuário consegue encontrar os dados sem maiores dificuldades, mesmo que estejam sem o PID;
- **F3:** como os arquivos de metadados e conjunto de dados estão normalmente separados, o identificador persistente precisa estar descrito nos metadados; e
- **F4:** para que os dados sejam encontrados é necessário que os metadados sejam indexados por mecanismos de busca, uma vez que somente ter os identificadores e os metadados não assegura a encontrabilidade dos dados. Por conseguinte, os usuários podem buscar

os elementos dos metadados através de um sistema que permite realizar pesquisas.

b) Acessíveis (*Accessible*) – questões sobre o uso de dados, consultas e reuso são determinadas para computador e usuários.

- **A1:** as autorizações de acesso e uso dos dados devem estar explicitadas, sem que seja necessária a utilização de comunicação ou ferramentas específicas;
 - **A1.1:** com o objetivo de potencializar a reutilização de dados, este ponto é capaz de interferir na escolha do repositório onde os dados serão depositados. Dessa maneira, o protocolo utilizado para conceder licenças deve, essencialmente, ser aberto, gratuito e capaz de ser aplicado por qualquer usuário; e
 - **A1.2:** um mecanismo que possibilite a autenticação e autorização para acesso aos dados precisa estar inserido no protocolo, em conformidade com as restrições de acesso a esses ativos e aos metadados.
- **A2:** mesmo após a degradação ou exclusão dos dados é necessário que seus metadados sejam preservados, fator que dependerá das estratégias de preservação envolvidas. Este ponto possibilita “a criação de índices para conjuntos de dados atuais e passados” (HENNING et al., 2019b, p. 178).

c) Interoperáveis (*Interoperable*) – os dados necessitam ser identificáveis para os computadores, de forma que possam ser associados com outros de forma automática.

- **I1:** para que os metadados possam ser representados é indicada a utilização de vocabulários controlados, ontologias e tesouros, além de ter uma estrutura bem estabelecida para a descrição dos metadados;
- **I2:** os vocabulários utilizados, também seguindo os princípios *FAIR*, precisam utilizar PID; e
- **I3:** este ponto refere-se a ser possível identificar conjuntos de dados que foram gerados a partir de outros conjuntos de dados, permitindo assim

que sejam associados por meio de ligações semânticas através dos padrões utilizados.

d) Reutilizáveis (*Reusable*) – com o objetivo de serem reutilizados e/ou combinados, dados e metadados necessitam estar consideravelmente descritos, de modo que computador e usuário possam reutilizá-los.

- **R1:** distingue-se do F2, pois aqui é voltado para os usuários. Deve fornecer metadados que possibilitem aos usuários identificar se os dados encontrados são adequados às suas necessidades. Sendo assim, é necessário que os metadados permitam não somente a descoberta, mas também a identificação de outros fatores envolvidos no contexto da geração dos dados;
 - **R1.1:** é necessário que esteja explícito quem pode ter acesso aos dados, para que fim poderá utilizá-los e sob quais condições, por meio de licenças de uso;
 - **R1.2:** é relevante não somente para identificar a origem, como também garantir crédito a quem gerou os dados. Os metadados devem fornecer elementos, como procedência dos dados; possíveis particularidades ou limitações; data de geração, condições do laboratório (se aplicável), identificar quem os compôs, como foram configurados os parâmetros, nome e versão de *softwares* utilizados (se aplicável); se são dados brutos ou processados; e por fim, qual versão dos dados se está disponibilizando. Todas essas informações precisam ser identificáveis por computador; e
 - **R1.3:** deve atender os padrões específicos das comunidades que constituem as áreas. Em vista disso, é essencial o uso de formatos de arquivo padronizados e a utilização de vocabulário que permita seu compartilhamento e reuso.

Para Henning et al. (2019b), cada letra da sigla *FAIR* representa um grupo de especificações que necessitam de conhecimento técnico e estruturais característicos. São princípios que devem ser entendidos como “um guia de preparação dos dados de pesquisa para o reuso, sob condições claramente

descritas por pessoas e máquinas” (IMMING, 2018, p.4, tradução nossa).

É importante ressaltar que o conceito de dados *FAIR* e dados abertos não devem ser confundidos, uma vez que para ser *FAIR* os dados não necessariamente precisam ser abertos, sendo conceitos complementares.

Dados *FAIR* podem e devem ser protegidos em alguns casos (ex.: dados referentes a patentes, levando em conta leis de propriedade intelectual e industrial que se divergem entre países, etc.), exigindo assim solicitação de acesso junto ao autor/instituição responsável (RODRIGUES, 2020, p. 36).

Em virtude de se fazer uma ciência mais aberta e eficiente, de modo que seja possível atender às solicitações de maneira mais assertiva, a Comissão Europeia publicou em 2016 um relatório intitulado *Realising the European Open Science Cloud*, da *Commission High Level Expert Group on the European Open Science Cloud* (EOSC), em que foram indicadas várias ações a fim de oferecer suporte à comunidade científica e agilizar os processos para que se comece a fazer uma ciência inovadora.

Conjuntamente com a Comissão Europeia, outras ações endossaram a adoção aos princípios *FAIR*, como a iniciativa *Big Data to Knowledge* (BD2K) de 2016 do *American National Institutes for Health* (NIH), na qual foi indicado que a Internet de Dados, essencialmente, empregue os princípios *FAIR*. De igual modo, na conferência *Jointly designing a Data FAIRPORT*, que ocorreu em 2014, a *Scientific Electronic Library Online* (SciELO) indicou que passaria a empregar os princípios *FAIR* na gestão de seus dados científicos. De fato, a *SciELO* atualmente, orienta seus usuários a utilizarem o *FAIR-Aware*¹¹ a fim de tornar os dados que serão depositados o mais *FAIR* possível:

A adoção de boas práticas na preparação de dados para depósito é um aspecto de fundamental importância para o acesso, compartilhamento e reuso dos dados de pesquisa, além de contribuir para que eles sejam *as FAIR as possible* (SCIELO, 2021, p. 2).

Apesar de os princípios *FAIR* já serem expostos e debatidos há algum tempo, ainda existem dificuldades em sua implementação, devido às interpretações que

¹¹ Ferramenta que auxilia usuários a tornarem seus dados científicos *FAIR*. Disponível em: <https://fairaware.dans.knaw.nl/>.

cada comunidade atribui a eles. Em decorrência do seu alto grau de complexidade e subjetividade, há algumas iniciativas se propondo a desenvolver ferramentas que serviriam para medir o grau de *FAIRness*¹² em ambientes digitais. Dentre essas iniciativas, destaca-se a *GO FAIR*.

Tendo seu início em 2018, o *Go FAIR* passou a oferecer um ecossistema aberto e inclusivo para indivíduos e instituições que trabalham através de redes de implementação, atuando em três pilares: *GO CHANGE*, *GO TRAIN* e *GO BUILD*. Cada uma dessas categorias atua em diferentes esferas, como sociocultural, treinamento de administradores e coordenação de tecnologias, para que seja possível não só a implementação dos princípios *FAIR*, como também sua manutenção e funcionamento de forma eficaz.

Atualmente, o Brasil é o único país na ibero-américa que aderiu à iniciativa. Com Sede no IBICT, possui a responsabilidade de difundir, apoiar e coordenar as atividades relacionadas à adoção da estratégia de implementação dos princípios *FAIR* definida pela iniciativa *GO FAIR* em todo o território brasileiro, o que inclui os principais repositórios de dados de pesquisa do país.

3.5 ATRIBUIÇÃO DE PALAVRAS-CHAVE COMO VALOR DO METADADO ASSUNTO

A indexação de assunto é “responsável por evidenciar os atributos temáticos dos documentos pela constituição de termos representativos de seus assuntos” (OLIVEIRA, 2017, p. 40). Porém, em uma definição mais ampla, a indexação de assuntos envolve a construção de índices, envolvendo processos mentais que se propõem a analisar o conteúdo informacional de registros de conhecimento e traduzi-los em linguagem compreensível por sistemas. Permite que os documentos já identificados sejam disponibilizados em bases de dados, catálogos etc. para recuperação da informação por meio de seus conteúdos, considerando procedimentos, normas e utilizando instrumentos que são capazes de guiar esse processo para que a disponibilização dos documentos ocorra com eficácia visando sua utilização pelos usuários.

As etapas do processo de indexação se caracterizam por:

¹² Nível *FAIR*, fazendo referência aos princípios *FAIR* como entidade.

1. Análise conceitual do conteúdo significativo do documento, ou seja, a identificação do assunto;
2. Expressão desta análise, através de um conjunto de palavras, frases ou códigos que representem o assunto;
3. Tradução das descrições dos assuntos relevantes para a linguagem de indexação;
4. Organização das descrições padronizadas dos assuntos de acordo com a sintaxe da linguagem de indexação (ROBREDO, 2005, p. 165 apud OLIVEIRA, 2017, p. 41).

Seu objetivo envolve “a intermediação entre usuário e documento no momento de busca e recuperação da informação em um sistema de informação” (PIOVEZAN, 2015, p. 28). Para que a intermediação ocorra de maneira assertiva e com o intuito de amenizar a subjetividade da indexação, se faz necessário o uso de normas que servem como guia para esse processo, dispostas nas diretrizes adotadas pelas instituições.

Considera-se ainda que

Para fins de indexação existem as linguagens documentárias/vocabulários controlados os quais são instrumentos que visam facilitar a comunicação mediante a padronização de termos para a descrição de conteúdo. Diante disso, verifica-se que tais instrumentos objetivam controlar as dispersões lexicais existentes nas linguagens naturais orientando também o indexador no que concerne à escolha dos termos para representar o assunto dos documentos, bem como indicar os melhores termos para que ele possa realizar a indexação do texto, atuando também como instrumento de comunicação entre a linguagem dos usuários e a linguagem do sistema (SANTOS, 2017, p. 10).

Em contrapartida, faz-se necessário que se estabeleça uma distinção entre indexação de assuntos e palavras-chave, uma vez que a atribuição de palavras-chave não necessita passar por processos mentais complexos, sendo caracterizada como “palavra representativa do conteúdo do documento, escolhida, preferentemente, em vocabulário controlado” (ABNT, 2021, p. 1). Dessa maneira, as palavras-chave podem ser extraídas de qualquer parte do documento, como título, resumo e até mesmo do texto, podendo ser atribuídas de forma livre “pelos autores do documento, por editores, e até ser geradas automaticamente” (GONÇALVES, 2008, p. 5), sem obrigatoriamente estarem atreladas a um vocabulário controlado.

Partindo do pressuposto que as palavras-chave apontam os principais conceitos de artigos científicos, são elementos essencialmente úteis para os

procedimentos de indexação.

Hartley e Kostoff (2003) elencam que as palavras-chave usualmente:

- Permitem que os leitores decidam se um artigo contém material relevante aos seus interesses;
- Fornecem aos leitores termos adequados para uso em pesquisas baseadas na *web* para localizar outros materiais do mesmo assunto ou em tópicos semelhantes;
- Ajudam os indexadores/editores a agrupar materiais relacionados, por exemplo as edições de final de ano de um determinado periódico ou um conjunto de anais de conferências;
- Permitem que editores/pesquisadores documentem mudanças em uma disciplina ao longo do tempo; e
- Vinculam às questões específicas de interesse a questões em níveis mais elevados de meta (abstração).

Ainda que as palavras-chave constituam uma parte essencial na recuperação da informação, nem todos os repositórios digitais disponibilizam em suas diretrizes aspectos relacionados à seleção dessas palavras. Em alguns casos, esse processo é realizado por profissionais que decidem quais palavras-chave melhor se adaptam a determinada pesquisa, processo esse que demanda alto desempenho financeiro e operacional da instituição. O que resulta disso é o próprio autor-pesquisador atribuindo quais palavras-chave representarão seu documento, fato que, sem a orientação de diretrizes que guiem esse processo, pode não ser o mais cabível por afetar a qualidade do metadado assunto e a recuperação da informação.

Em um repositório de dados de pesquisa

Os termos são pontos de acesso para os documentos, haja vista que é possível recuperar esses objetos informacionais sob qualquer uma das suas palavras-chave (que foram empregadas no sistema no ato da sua representação) ou até mesmo por meio da combinação delas” (SANTOS, 2017, p. 6).

Mediante as palavras-chave atribuídas é possível identificar e acessar os documentos buscados. Desta forma, esse processo que constitui uma parte da

representação temática da informação faz conexão entre usuário e documento.

Caplan (2003) define metadados como dados que são usados para descrever outros dados, afirmando ainda que devem descrever um recurso informacional. Neste mesmo sentido, Greenberg (2005) define metadados como dados estruturados sobre um objeto que suporta funções associadas ao objeto designado, com um objeto sendo qualquer entidade, forma ou modo para o qual os dados contextuais podem ser registrados. Já Smiraglia (2005) afirma que os metadados são dados provenientes diretamente de um recurso informacional que mais tarde é reformulado conforme o esquema de metadados, de modo que possa criar uma representação do recurso. Nesse contexto, a relação entre os metadados e os objetos que os dados representam ou substituem é direta, uma vez que a padronização desses metadados visa possibilitar o acesso e a recuperação de documentos considerados relevantes para o usuário.

Em vista das diversas aplicações dos metadados, surgem várias tipologias que se referem ao uso funcional e a intenção do metadado, sendo classificados em descritivos, administrativos e estruturais (CAPLAN, 2003). Suas funções envolvem a descoberta de recursos, recuperação de informação, gerenciamento de recursos, uso de recursos, uso de recursos pelo público apropriado, autenticação de recursos e outras atividades relacionadas à proveniência, recursos linkados com recursos relacionados e recurso de *hardware* e necessidade de *software*.

Antes de qualquer coisa, os metadados necessitam atender às especificações dos princípios *FAIR*, observando principalmente que sejam localizáveis e interoperáveis, consideradas as mais significativas entre elas. Contudo, a interoperabilidade somente é possível mediante os padrões de metadados.

Conforme exposto por Hermann (2017), padrão de metadados é o conjunto padrão de campos obrigatórios e opcionais que descrevem os conjuntos de dados e tem por objetivo que os dados se apresentem de uma maneira clara e uniforme para os usuários. Podem ser também “estruturas formais de descrição de recursos para comunidades específicas, de modo a apoiar a gestão e a preservação de objetos digitais em ambientes informacionais” (FORMENTON et. al., 2017).

A implementação desses padrões em ambientes digitais depende da avaliação dos aspectos a serem considerados no desenvolvimento de projetos de implantação de bibliotecas digitais. Foulonneau e Riley (2008) afirmam que os fatores para avaliação podem ser definidos como: a instituição, o padrão, os

materiais e o projeto. Para Ikematu (2001), organizações no nível mais baixo da hierarquia gerenciam dados brutos. Organizações mais avançadas são capazes de administrar seus recursos de informação no nível de Informação, Conhecimento ou Sabedoria.

Para Riley (2009), os padrões de metadados são divididos em: domínio, comunidade, função e propósito, em que:

- O **domínio** é referente aos tipos de materiais com os quais o padrão se destina a ser usado ou para os quais poderia ser útil. As categorias específicas representadas aqui não pretendem ser exaustivas, nem são mutuamente exclusivas. Em vez disso, elas se concentram em alguns tipos de materiais comuns que são gerenciados pelo patrimônio cultural e outras organizações de informação. Exemplo: objetos culturais, conjuntos de dados, dados geoespaciais, imagens em movimento, materiais musicais, textos acadêmicos e recursos visuais.
- A categoria de **comunidade** abarca os grupos que atualmente usam o padrão potencialmente. Aqueles que originaram um padrão ou que são os públicos primários são correspondências mais fortes, enquanto aqueles que poderiam usar o padrão de forma eficaz não o fazem frequentemente, são correspondências mais fracas. Exemplo: Bibliotecas, arquivos, museus e indústria da informação.
- A **função** seria o papel que um padrão desempenha na criação e armazenamento de metadados. Algumas funções definem as entidades básicas a serem descritas, outras definem campos específicos, outras fornecem orientações sobre como registrar um elemento de dados específico e ainda, outras definem estruturas de dados concretas para o armazenamento de informações. Exemplo: modelos conceituais, padrões de conteúdo, vocabulários controlados, *frameworks* ou tecnologias, linguagens de marcação, formato de registros e padrões de estruturas.
- A última categoria, sendo o **propósito**, se refere ao tipo geral que o padrão de metadados foi projetado para registrar. Normalmente, um padrão será fortemente focado em um propósito, mas incluirá alguns elementos de dados para outros propósitos considerados especialmente importantes. Exemplo: dados, metadados descritivos, *wrappers* de metadados,

metadados de preservação, metadados de direitos, metadados estruturais e metadados técnicos.

Quando se tem um conjunto de elementos de metadados, tem-se então um esquema de metadados que são constituídos pela semântica, conteúdo e linguagem. Neste sentido, Greenberg (2005) aponta que esquema de metadado é:

- Uma coleção de elementos de metadados reunidos para suportar uma função, ou uma série de funções para um objeto de informação;
- Uma coleção de elementos de metadados, formando um contêiner estruturado, aos quais os valores dos dados são adicionados. Os valores dos dados podem ser descontrolados ou controlado; e
- Uma coleção de elementos de dados, com seus atributos formalizados em uma especificação (ou um dicionário de dados).

Considera-se que os metadados e os processos que os envolvem auxiliam na organização e recuperação do conhecimento científico para mais tarde ser utilizado pelos usuários da informação. Em vista disso, a representação desses metadados precisa ser realizada corretamente, pois quando há erro na representação, haverá erro na recuperação dessas informações e, conseqüentemente, na comunicação entre usuário e sistema. Dessa forma, no âmbito dos repositórios de dados de pesquisa, comumente são empregados metadados descritivos que são utilizados para fins de descoberta, identificação e seleção, podendo ser úteis também para a colocação e aquisição.

O metadado assunto ou temático contempla as informações dedicadas à representação do conteúdo dos recursos informacionais incluídas no momento do depósito dos dados de pesquisa no repositório, sendo feita pelo próprio autor-pesquisador (indexação por extração) ou por um profissional bibliotecário. Os assuntos expressos por palavras-chave representam o conteúdo dos dados de pesquisa, sendo os termos empregados com o propósito de representação e recuperação por assunto.

Para Rodrigues et. al. (2010), ainda há um longo caminho a percorrer quanto ao estabelecimento de práticas de descrição adequadas e para a normalização dos descritores que serão usados em repositórios de dados de pesquisa. Segundo o

autor, “a descrição normalizada dificilmente poderá cobrir toda a especificidade de um conjunto de dados, mas muitos dados descritivos a nível de conjunto ganharão em ser normalizados” (RODRIGUES et. al. 2010, p. 28).

A função da representação temática da informação contribui com melhorias nos procedimentos e viabiliza a recuperação por assunto, resultando no atendimento às necessidades e demandas dos usuários em relação aos produtos e serviços oferecidos pelos sistemas. O estabelecimento de considerações sobre o processo de atribuição de palavras-chave para autores-pesquisadores responsáveis pela inclusão dos dados de pesquisa nos repositórios de dados de pesquisa favorece uma representação da informação mais adequada e compatível com os preceitos propostos pela indexação de assunto e considerados na prática de bibliotecários-indexadores.

Na maioria dos repositórios digitais que não adotam o autoarquivamento, a representação por assunto é conduzida pelos bibliotecários-indexadores, os quais possuem treinamentos e formação adequados para a condução da atividade. Contudo, em repositórios digitais que recorrem ao autoarquivamento enquanto mecanismo que permite ao autor-pesquisador realizar o depósito de suas publicações, dados e outros materiais acadêmicos, a inexistência de diretrizes e orientações específicas para a representação do conteúdo por parte do autor-pesquisador na condição de autor-indexador mostra-se um cenário preocupante.

Em vista disso, reitera-se a relevância de estudos que explorem esta temática em repositórios de dados de pesquisa, ao passo que a falta de procedimentos adequados para a representação do conteúdo dos dados de pesquisa impacta no compartilhamento e acesso desses recursos informacionais, inviabilizando a encontrabilidade da informação e a reprodutibilidade científica.

4 ANÁLISE E DISCUSSÃO DOS RESULTADOS

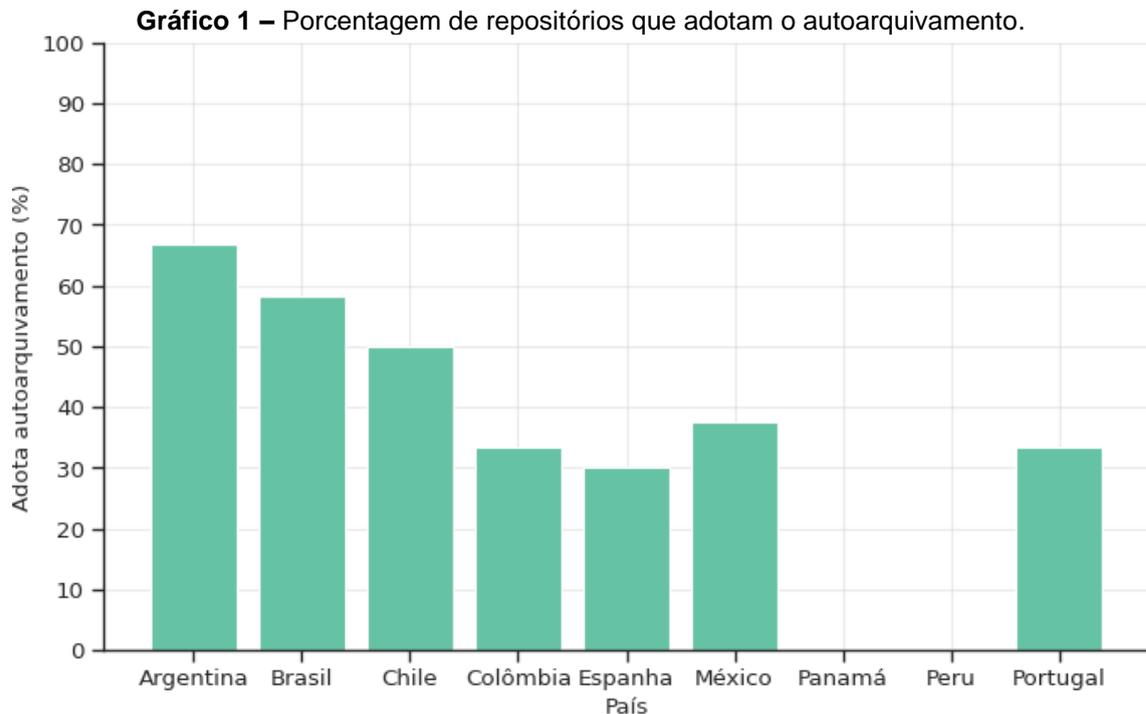
Como visto no decorrer da presente pesquisa, há vários caminhos pelos quais o movimento de acesso aberto pode se permear, contando com inúmeras iniciativas que endossam suas práticas em prol de que se faça uma ciência mais aberta. Dentre as iniciativas e ferramentas utilizadas figuram os repositórios de dados de pesquisa que auxiliam nas ações voltadas para dados originados de pesquisas científicas, preservando, curando e compartilhando. São considerados fundamentais para a propagação da Ciência Aberta, uma vez que os dados depositados nesses ambientes se encontram disponíveis para uso de qualquer usuário que consiga acessá-los.

Outra iniciativa que contribui com os preceitos da Ciência Aberta é a prática do autoarquivamento adotada por repositórios digitais, e definida pela BOAI como estratégia para permitir aos autores-pesquisadores tornarem seus estudos disponíveis na *web* de forma mais eficaz. Em decorrência da adoção dos princípios *FAIR* por repositórios de dados de pesquisa, torna-se possível encontrar todos os dados relacionados a uma mesma pesquisa, otimizando o tempo dos pesquisadores na reprodutibilidade científica e na produção de novos conhecimentos.

Diante deste cenário, surge a preocupação em observar como os repositórios de dados de pesquisa que adotam o autoarquivamento estão orientando seus autores-pesquisadores quanto ao preenchimento dos metadados, especificamente do metadado assunto, no momento do depósito dos seus documentos. Para tanto, foram feitas análises em 26 repositórios de dados de pesquisa pertencentes aos países da ibero-américa que adotam o autoarquivamento, a partir das diretrizes disponíveis e dos próprios ambientes desses repositórios. Da mesma forma, foram estabelecidos indicadores para auxiliar nas análises dos achados da pesquisa, de forma que fosse possível identificar quais diretrizes estão sendo determinadas para a atribuição de palavras-chave.

Como mencionado anteriormente, de todos os países que compoem a ibero-américa, apenas 9 possuem repositórios de dados de pesquisa registrados no *Re3data*, sendo que nem todos os repositórios listados praticam o autoarquivamento. No Gráfico 1, a Argentina aparece com 67% de seus repositórios de dados de pesquisa adotando o autoarquivamento, seguida por Brasil e Chile com 58% e 50%, respectivamente.

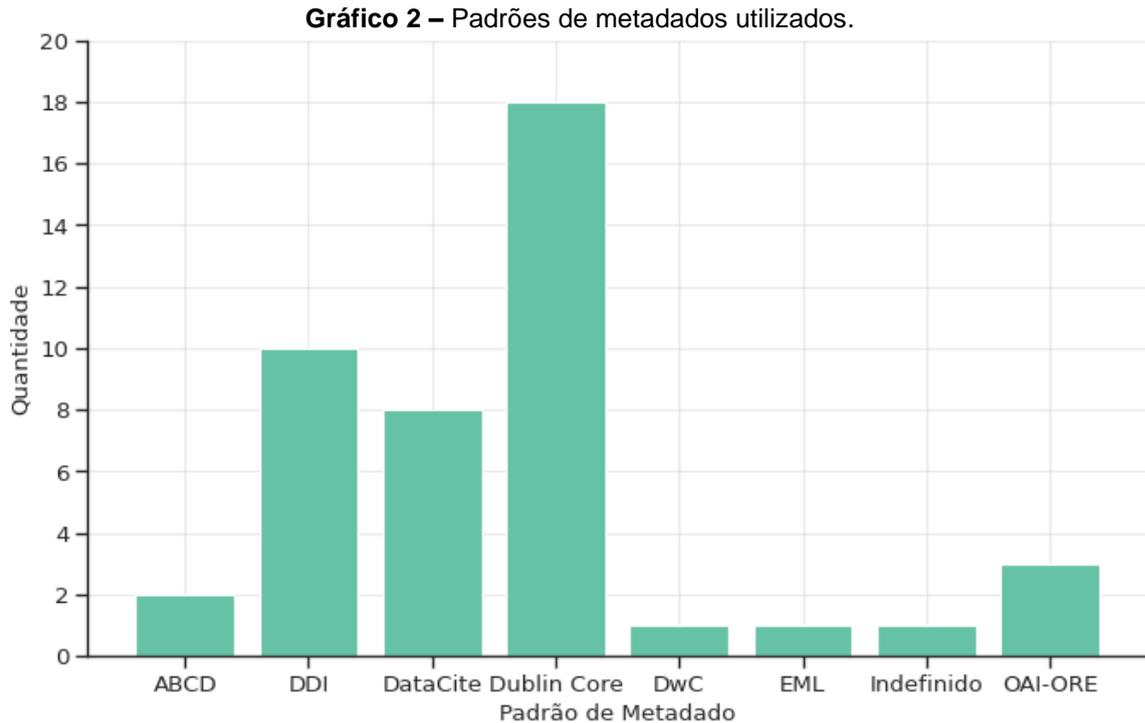
Os países com menos repositórios de dados de pesquisa que adotam o autoarquivamento são Panamá, Peru e Espanha, com este último tendo apenas 30%, mesmo sendo o país ibero-americano com o maior número de repositórios de dados de pesquisa registrados no *Re3data*.



Fonte: Dados de pesquisa.

Identificou-se ainda 7 padrões de metadados adotados por esses repositórios de dados de pesquisa, observados no Gráfico 2: *Access to Biological Collection Data* (ABCD), *Darwin Core*, *Ecological Metadata Language* (EML), *Open Archives Initiative Object Reuse and Exchange* (OAI-ORE). O Dublin Core, o *Data Documentation Initiative Alliance* (DDI Alliance ou somente DDI) e o *DataCite Metadata Schema* sendo os mais utilizados.

Alguns repositórios utilizam mais de um padrão de metadados, como é o caso do Repositório de dados de pesquisa do Instituto Federal Goiano - Campus Urutaí, do *Repositorio de datos de investigacion de la Universidad del Rosario*, *RiuNet*, entre outros. Não foi possível identificar o padrão de metadado utilizado apenas por um repositório (ANEXO A).



Fonte: Dados de pesquisa.

O *Access to Biological Collection Data* (ABCD)¹³ é um padrão utilizado para descrever dados primários e de biodiversidade, como espécimes e observações realizadas, sendo empregado em coleções biológicas. Assim como o ABCD, o *Darwin Core*¹⁴ também é voltado para a diversidade biológica, mantido pelo *Darwin Core Maintenance Interest Group*.

O DDI¹⁵ surgiu em 2003, sendo uma colaboração internacional com o intuito de auxiliar na definição de padrões de metadados e ferramentas semânticas que descrevessem dados voltados para a área das ciências sociais, que abarcassem atividades humanas e outros dados originados de métodos observacionais. Por pouco mais de 20 anos a DDI tem trabalhado para facilitar a descrição e documentação de dados de forma que atendesse as demandas da comunidade científica que atua nas ciências sociais, em atenção aos requisitos mundiais de padrões de metadados.

O *DataCite Metadata Schema*¹⁶, que despontou em 2009, é uma organização global sem fins lucrativos que atua fornecendo Identificadores Persistentes para

¹³ Disponível em: <https://abcd.tdwg.org/>

¹⁴ Disponível em: <https://dwc.tdwg.org/>

¹⁵ Disponível em: <https://ddialliance.org/>.

¹⁶ Disponível em: <https://schema.datacite.org/>.

dados de pesquisa e outros resultados que possam envolver documentos provenientes de pesquisas. Uma vez que um Identificador Persistente é atribuído, suas saídas se tornam identificáveis, o que torna possível que os metadados associados sejam disponibilizados para a comunidade científica.

A *Dublin Core Metadata Initiative*¹⁷ – DCMI é uma instituição voltada para apoiar o desenvolvimento do *design* de metadados, assim como as práticas que envolvem todas as ações voltadas para o tema. Com início de suas atividades no ano de 1995, em Dublin, cidade localizada no estado norte-americano de Ohio, onde ocorreu o evento denominado Oficina de Metadados da OCLC/NCSA, no qual se discutiu um conjunto básico de semântica para recursos baseados na *web*. Esta ação se mostrou essencial, pois permitiu que fosse possível categorizar a *web*, o que auxiliaria na pesquisa e recuperação de informações. Em vista disso, esse padrão de metadado funciona de forma aberta e é suportado por um modelo que inclui associações pagas.

Quanto ao restante dos padrões de metadados, o *Ecological Metadata Language* (EML)¹⁸ é empregado na descrição de dados originados da geociência e ciências ambientais, possuindo módulos capazes de identificar e citar pacotes de dados, descrever a extensão espacial, temporal, taxonômica e temática dos dados, descrever métodos e protocolos de pesquisa, descrever a estrutura e o conteúdo dos dados dentro de pacotes de dados às vezes complexos e para descrever dados com vocabulários semânticos.

Por fim, o *Open Archives Initiative Object Reuse and Exchange* (OAI-ORE)¹⁹ utiliza um mapa de recursos que torna possível a associação entre uma identidade e agregações (também conhecidas como objetos digitais compostos) de recursos permitindo a identificação de sua estrutura e semântica.

Ratifica-se a relevância da pesquisa para a Ciência da Informação, sobretudo para a área de Organização do Conhecimento, posto que a atribuição de palavras-chave no autoarquivamento atua como a indexação de assuntos, sendo fundamental para que a recuperação da informação ocorra de maneira eficaz. Desta forma, o trabalho tenciona contribuir com os estudos direcionados às diretrizes estabelecidas por repositórios de dados de pesquisa, assim como destacar a relevância do

¹⁷ Disponível em: <https://www.dublincore.org/>.

¹⁸ Disponível em: <https://eml.ecoinformatics.org/>

¹⁹ Disponível em: <https://www.openarchives.org/ore/>

metadado assunto para a recuperação da informação, contribuindo com os preceitos estabelecidos pela Ciência Aberta.

A seguir, inicia-se a apresentação das análises conforme os indicadores definidos para a pesquisa e descritos na Tabela 2, quais sejam: Diretrizes; Metadado Assunto; Palavras-chave; Preenchimento do metadado assunto/palavra-chave; Linguagem documentária; e Validação das palavras-chave por bibliotecários.

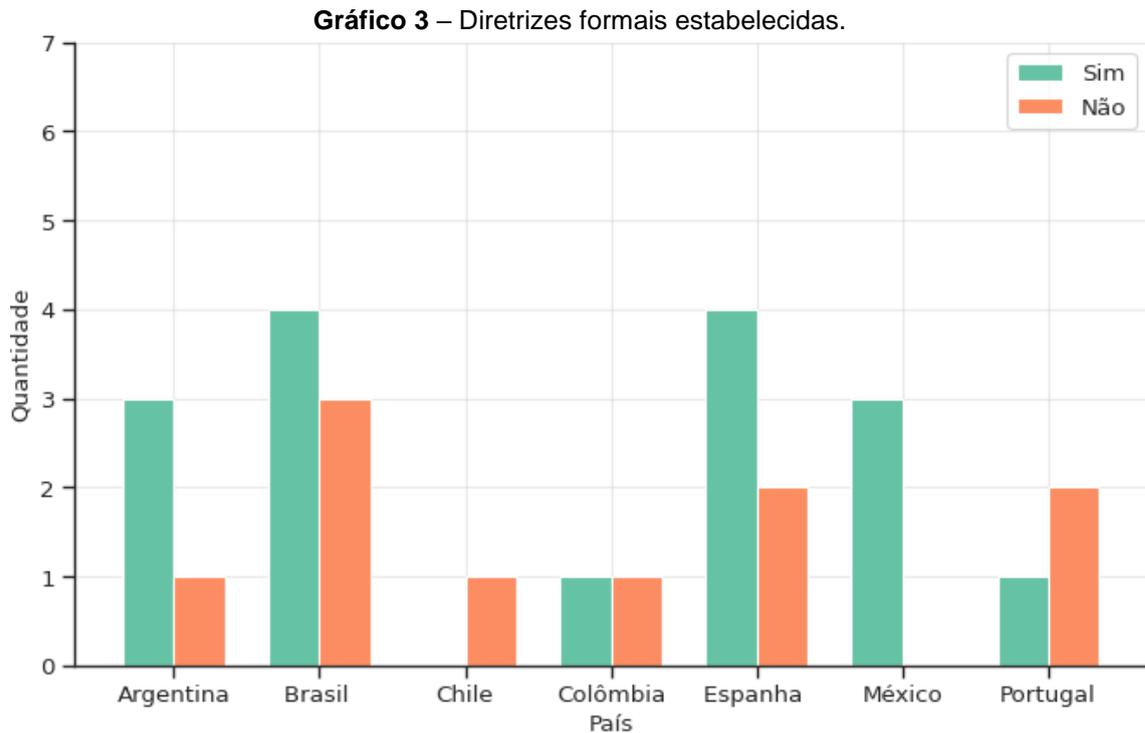
4.1 DIRETRIZES

As diretrizes de repositórios de dados de pesquisa se constituem em documentos que tem a função de guiar os processos e tomadas de decisão relacionados ao funcionamento dos repositórios. Nesses documentos, podem figurar indicações específicas para a atribuição de palavras-chave pelos autores-pesquisadores no autoarquivamento, ou ainda serem diretrizes específicas para o autoarquivamento.

Uma vez que as diretrizes estejam estabelecidas em repositórios de dados de pesquisa que adotam o autoarquivamento e disponíveis para consulta pelos autores-pesquisadores, as inconsistências decorrentes do processo de atribuição de palavras-chave tendem a serem minimizadas, resultando em uma representação de conteúdos dos dados de pesquisa mais precisa e que favoreça uma recuperação de informações verdadeiramente relevante para os usuários do sistema.

Diante deste cenário, este primeiro indicador decorre da necessidade de se observar se os repositórios de dados de pesquisa ibero-americanos que adotam o autoarquivamento possuem diretrizes formais disponíveis para consulta e, se nesses documentos, existem instruções aos autores-pesquisadores no momento do depósito de seus documentos. Para tanto, os ambientes dos repositórios de dados de pesquisa identificados foram examinados de forma pormenorizada com intuito de identificar as suas respectivas diretrizes e/ou outros documentos de orientação. O *Re3data* possui uma aba na qual disponibiliza alguns documentos relacionados a cada repositório de dados de pesquisa incluído na ferramenta, que também foram considerados, em alguns casos, para a análise desse indicador quando os repositórios não possuíam diretrizes formais estabelecidas.

Desta forma, revelou-se o cenário indicado no Gráfico 3 para o indicador sobre diretrizes formais.



Fonte: Dados de pesquisa.

Dos 26 repositórios de dados de pesquisa analisados, 16 possuem documentos formais que se encontram disponibilizados em seus ambientes, sendo Brasil e Espanha os países com o maior número de repositórios de dados de pesquisa com diretrizes publicadas e disponíveis para consulta. Portugal e Chile são os países que possuem o menor número de repositórios de dados de pesquisa com diretrizes estabelecidas, contudo vale ressaltar que o Chile só possui um repositório de dados de pesquisa que adota o autoarquivamento, até o momento.

O fato de mais da metade dos sistemas analisados terem diretrizes estabelecidas contribui para que as ações relacionadas a esses repositórios de dados de pesquisa sejam padronizadas, ao mesmo tempo em que se encontram ao alcance do usuário. Entretanto, embora esses documentos estejam estabelecidos, nem sempre são identificados por nomes adequados de forma a deixar claro para o usuário sobre o seu conteúdo, como é o caso do Globe (Anexo B).

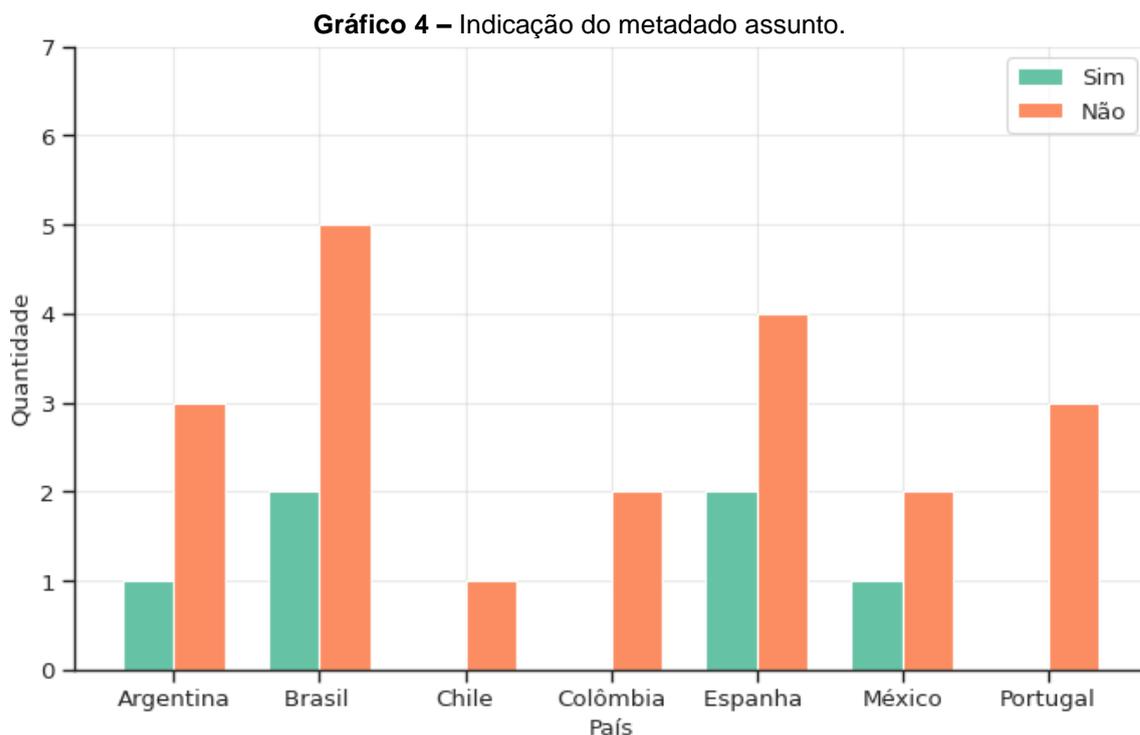
Outro fator a ser observado é que muitos dos repositórios de dados de pesquisa considerados nesta pesquisa não possuem diretrizes estabelecidas e fazem uso dos manuais dos *softwares* como DSpace e Dataverse para orientarem os seus usuários. Considerando que nesses manuais estão contidas diretrizes relacionadas a todo tipo de questões envolvendo os *softwares*, a leitura desses

manuais, alguns com 900 páginas (caso do DSpace), pode se tornar uma dificuldade para o usuário que esteja em busca de orientações específicas, como no caso do autor-pesquisador no momento da atribuição de palavras-chave.

4.2 METADADO ASSUNTO

Os metadados de assunto desempenham um papel de relevância em repositórios digitais, atuando diretamente na recuperação da informação, sendo “[...] considerados elementos que visam proporcionar o desenvolvimento de sistemas de recuperação de informação, permitindo a interoperabilidade entre eles e o acesso mais amplo às informações disponibilizadas em diversos ambientes” (SANTOS; NEVES, 2018, p. 1010).

Observa-se no Gráfico 4 que poucos repositórios de dados de pesquisa fazem indicação especificamente ao termo ‘metadado assunto’. Somente os repositórios *Humadoc*, *SciELOData*, *CEDAP*, *IoChem – BD*, *UPCommons* e *ciRcle*, da Argentina, Brasil, Espanha e México possuem em suas diretrizes pontos em que empregam essa terminologia. A causa para este fato reside em utilizar-se nomes que diferem entre si para representar o campo em que se insere as informações a respeito do assunto dos dados.



Fonte: dados de pesquisa.

Essa falta de padronização pode vir a ter um efeito negativo para a representação da informação, uma vez que foi identificado que, em alguns documentos, não é indicado nem mesmo qual o campo que o autor-pesquisador deve preencher para indicar o assunto de seus dados. Para além desta falta, não fica claro em nenhum desses documentos, nem nos ambientes dos próprios repositórios de dados de pesquisa investigados, de que forma a representação de conteúdo dos dados de pesquisa acontece, deixando os usuários sem maiores informações.

4.3 PALAVRAS-CHAVE

As palavras-chave são termos empregados com a função de descrever o conteúdo dos dados de pesquisa, também designadas como descritores. É possível atribuí-las por meio do metadado assunto, embora na maioria das diretrizes dos repositórios de dados de pesquisa analisados este campo aparece, simplesmente, como 'palavras-chave'.

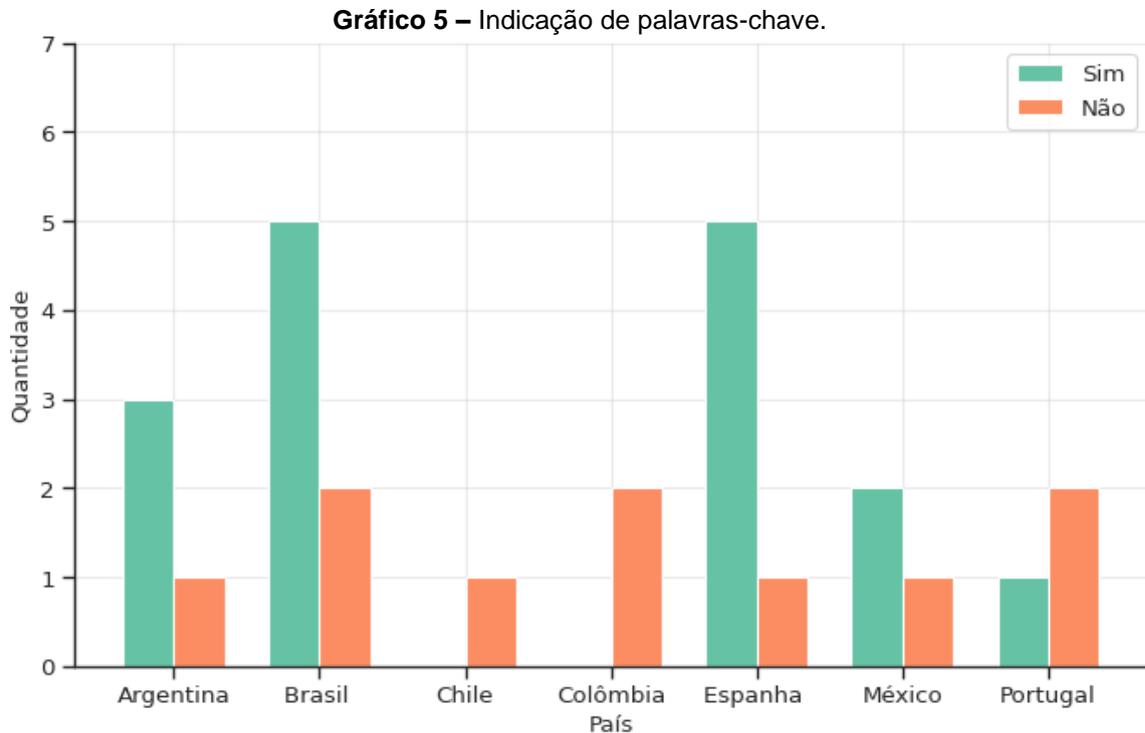
A atribuição de palavras-chave pelo próprio autor-pesquisador durante o autoarquivamento é uma prática recente. Discussões teóricas sobre esse tema são recentes. Além disso, o estabelecimento de diretrizes para a condução dessa prática em repositórios digitais que adotam o autoarquivamento ou são recentes, ou estão em discussão e desenvolvimento.

Em vista disso,

A relevância de palavras-chave para tais aplicações reside no fato de que é o próprio autor quem garante a representatividade "chave" de seus textos. Isso é compreensível, porque o autor produziu o texto, é o especialista do tema tratado e tem domínio do conteúdo e do vocabulário utilizado. Portanto, é ideal que ele próprio atribua palavras-chave que representem o texto por ele produzido (FUJITA; TARTAROTTI, 2020, p. 333).

Observou-se no Gráfico 5 que do total de repositórios de dados de pesquisa investigados, pouco mais da metade (16 repositórios) utilizam o termo 'palavras-chave' em suas diretrizes. Quanto ao restante dos repositórios, ou indicam a representação de conteúdo de dados de pesquisa pelo termo 'assunto' ou não a

indicam de forma alguma. Há ainda poucos documentos que fazem alusão aos dois termos, indicando que as palavras-chave estão contidas no metadado assunto.



Fonte: dados de pesquisa.

Fato é que, ao se tornar um indexador, como afirmado por Fujita e Tartarotti (2020), o autor-pesquisador precisa deixar de enxergar as palavras-chave como mecanismos que simplesmente descrevem seus dados de pesquisa e passar a vê-las desempenhando outras funções para que seja possível a utilização de padrões que abarquem no futuro outras ações envolvendo seus dados.

4.4 PREENCHIMENTO DO METADADO ASSUNTO/PALAVRAS-CHAVE

O preenchimento do metadado assunto por meio da atribuição de palavras-chave mostra-se essencial em diversos aspectos da representação da informação. Não somente serve para os propósitos da recuperação dos dados de pesquisa, mas como também fornece visibilidade para eles no meio digital. Portanto, empregar as palavras-chave adequadas trará visibilidade aos dados de pesquisa, como também ao autor e até mesmo a instituição a qual ele está associado.

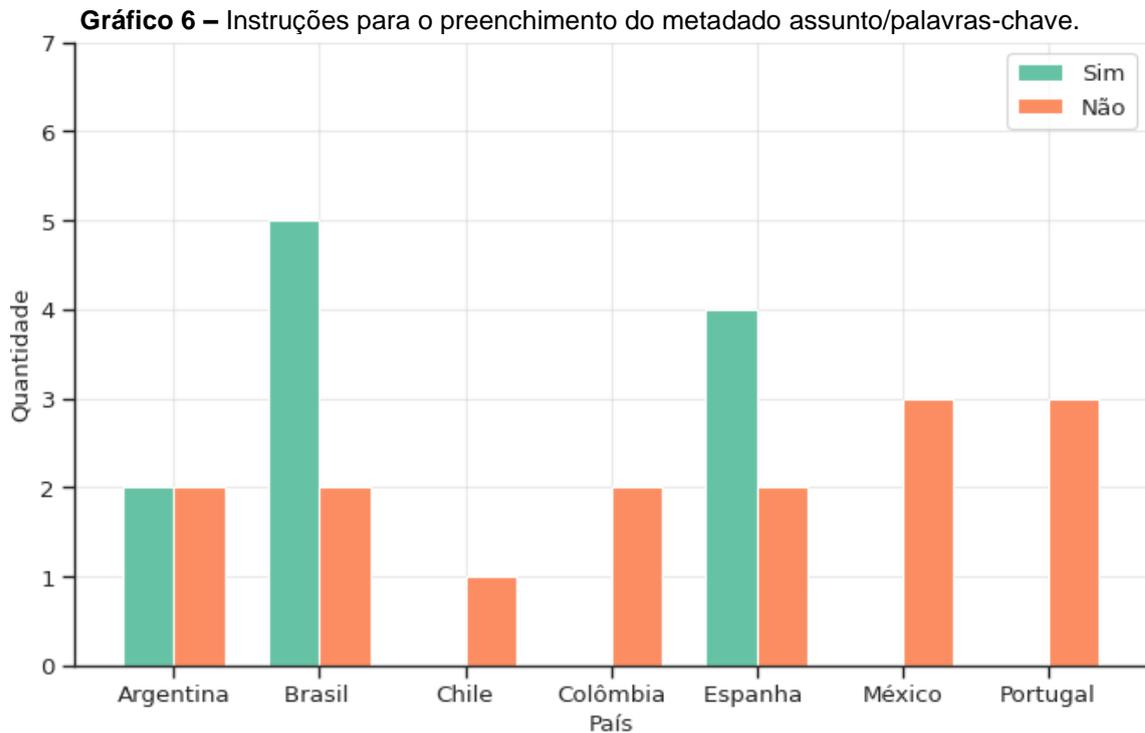
Embora estabelecer padrões uniformes pareça ser uma limitação, no caso do emprego das palavras-chave torna-se uma necessidade estratégica primordial, dado

que a recuperação dos dados de pesquisa depende da atribuição adequada no momento do depósito. Este fator traz ainda uma questão mais preocupante, pois ao não utilizar palavras-chave que sejam consistentes, o potencial acadêmico e científico da área se torna disperso.

Em vista disso, definir algumas estratégias para o emprego dos descritores mostra-se benéfico, pois até mesmo a posição das palavras-chave é relevante para a busca dos dados de pesquisa, sendo necessário que primeiro se coloquem os termos que fazem parte do tema central dos dados, seguido pelos termos dos assuntos secundários, criando assim uma ordem decrescente. Quando são usados termos isolados, que podem ter vários significados, como o termo 'organização', por exemplo, a busca será mais genérica e pode não retornar aquilo que o usuário de fato deseja encontrar. Porém, quando os termos são associados a outros termos que oferecem um tópico mais bem definido, os resultados de busca se tornam mais específicos.

Quanto às orientações para atribuição das palavras-chave nos repositórios de dados de pesquisa identificados, o cenário se mostrou preocupante, em virtude de menos da metade do total dos repositórios de dados de pesquisa contemplados incluírem em suas diretrizes ou em seus ambientes digitais instruções claras para o preenchimento do metadado assunto. Apenas 11 repositórios de dados de pesquisa ibero-americanos que adotam o autoarquivamento estabelecem essas orientações, sendo o *Portal de datos de Biodiversidad*, *Humadoc repositorio*, *Globe*, *PPBio Data Repository*, *ScieloData*, *CEDAP*, *Repositório Institucional UNESP*, *IoChem – BD*, *UPCommons*, *Digibug* e *RiuNet*, pertencentes a Argentina, Brasil e Espanha, conforme indicado no Gráfico 6.

Observou-se que os repositórios de dados de pesquisa que continham tais orientações em seus documentos costumam elencar cada campo a ser preenchido, seguido por instruções claras de como inserir as palavras-chave individualmente e sem pontuação no final, o uso ou não uso de caixa alta, e recomendações para o uso de linguagens documentárias, podendo ser disponibilizadas pelo próprio repositório ou com indicações de onde podem ser acessadas. Nesse último caso, cabe destacar que “o autor, quando atribui palavras-chaves, preenche um metadado, sem orientação ou auxílio profissional, que será preservado quanto à padronizações ou qualquer proposta de controle de vocabulário, após preenchimento” (FUJITA; TARTAROTTI, 2020, p. 335).



Fonte: dados de pesquisa.

Neste sentido, a disponibilização de diretrizes como guias de auto-depósito ou autoarquivamento é uma prática de gestão dos repositórios de dados de pesquisa considerada primordial para guiar os autores-pesquisadores quanto aos passos necessários para um preenchimento correto do metadado assunto.

4.5 LINGUAGEM DOCUMENTÁRIA

Segundo Lara (2004), as linguagens documentárias são ferramentas que traduzem a linguagem comum em conceitos a fim de que estes sejam reconhecidos por sistemas de informação e auxiliem na mediação entre sistemas e usuários. A fim de que essa comunicação ocorra de forma efetiva, a autora apresenta ainda algumas características inerentes às linguagens documentárias, constituindo-se na questão de que necessitam atuar como uma linguagem compreensível, que seja base para a interpretação do sentido, apresentar-se como metalinguagem e ainda, ser capaz de considerar o usuário como parte necessária aos seus processos.

Com o objetivo de possibilitar a tradução dos conceitos representativos dos dados de pesquisa em termos precisos, as linguagens documentárias se fazem essenciais, operando como base nos processos de organização, representação e

recuperação da informação. Desta forma, considera-se relevante a definição de estratégias para o emprego das palavras-chaves “objetivando-se controlar sinônimos, diferenciar homógrafos e ligar termos por meio de relações hierárquicas, associativas e de equivalência (ou seja, utiliza-se uma linguagem artificial/linguagem controlada) (SANTOS, 2017, p. 7), pois o uso de linguagem natural sem o auxílio de ferramentas desenvolvidas para apoiar o processo de atribuir as palavras-chave traz consequências negativas para as práticas de representação da informação.

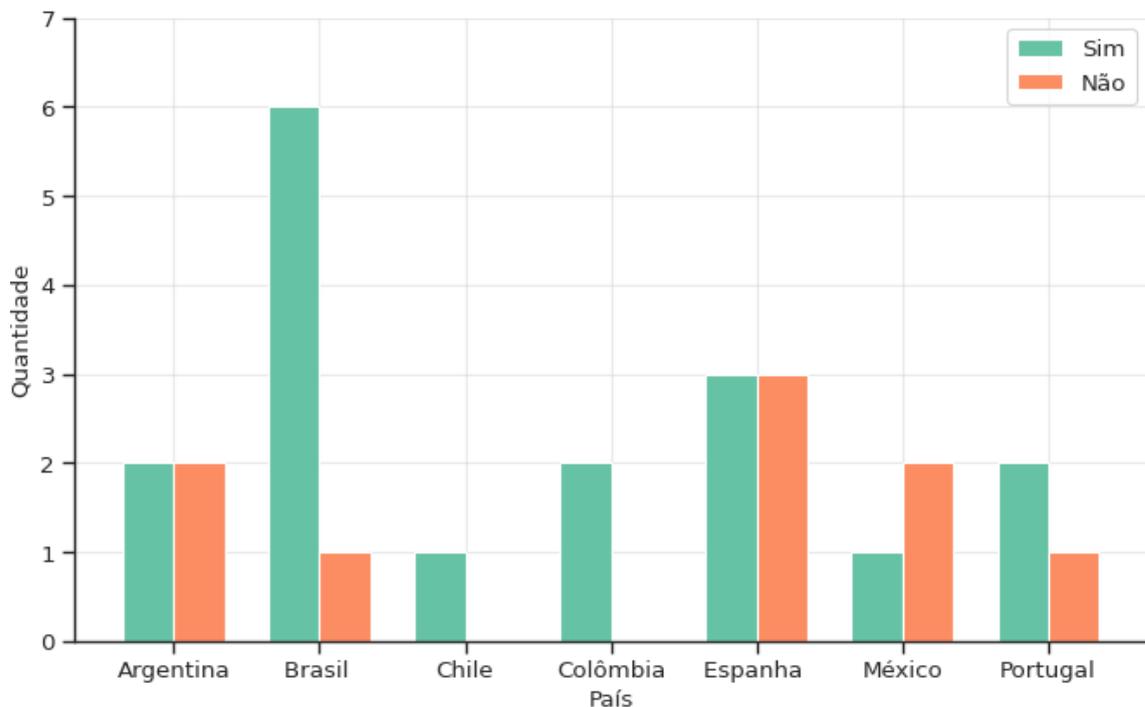
O fato de o autor-pesquisador especialista de um domínio do conhecimento atuar como autor-indexador leva o uso da linguagem natural para os ambientes dos repositórios digitais, de modo que se difere do uso das linguagens documentárias que são originalmente desenvolvidas para atender as necessidades dos usuários. Mesmo que os autores-pesquisadores utilizem como descritores termos que acreditam que fazem sentido para a descrição de um conteúdo, a realidade pode ser diferente, tendo significado apenas para quem os empregou, de forma que as comunidades ou usuários dos repositórios não conseguirão recuperar as informações por não estarem familiarizados com os termos utilizados. Assim, a falta de padronização acarreta na dissociação das práticas da representação temática da informação comumente conduzidas por bibliotecários-indexadores.

Algumas complicações para sistemas que utilizam linguagem natural para a atribuição de palavras-chave sem que seja indicado o uso de quaisquer linguagens documentárias foram identificadas, sendo dispostas a seguir:

- 1) **Plurais:** podem prejudicar um sistema de marcação. Por exemplo, se as tags gato e gatos forem distintas, uma consulta para uma não recuperará ambas, a menos que o sistema de pesquisa inteligente tenha a capacidade de realizar tais substituições incorporadas;
- 2) **Polissemia:** refere-se a uma palavra que tem dois ou mais significados semelhantes.
- 3) **Sinonímia:** palavras diferentes com significados semelhantes ou idênticos, apresenta um problema maior para os sistemas de marcação porque a inconsistência entre os termos usados na marcação pode tornar muito difícil para um pesquisador ter certeza de que todos os itens relevantes foram encontrados.
- 4) **Profundidade (especificidade) da marcação:** significa quão específico deve ser o usuário (indexador) ao traduzir um conceito em descritores. Os recursos da *Web* podem ser marcados em vários níveis de especificidade, desde assuntos muito amplos, desde o título e resumo até o nível do parágrafo. A profundidade das tags refere-se a quantas tags existem em relação a um recurso da *Web* no sistema (NORUZI, 2007, n. p, grifo nosso).

No Gráfico 7, observa-se que mais da metade dos repositórios de dados de pesquisa recomendam o uso de linguagens documentárias, como vocabulários controlados e tesouros, para o emprego dos descritores. Esse fator contribui para atenuar escolhas inadequadas de descritores, minimizando o uso de homógrafos, que apresentam formas similares de escrita, porém com significados distintos, assim como os erros ortográficos e descritores regionalizados. Auxilia na padronização dos termos, com definições claras dos descritores em maiúsculos e minúsculos, como também descritores dissonantes.

Gráfico 7 – Indicação de linguagens documentárias.



Fonte: dados de pesquisa.

Diante do cenário exposto, acredita-se que o uso de linguagens documentárias auxilia não somente na padronização dos termos, mas também na comunicação entre usuário/usuário quanto da representação dos conteúdos. Para Santos (2017), as linguagens documentárias têm por finalidade conter as dispersões lexicais originadas do uso de linguagem natural, ao passo que guia o autor-indexador no processo de seleção das palavras-chave para a representação do conteúdo dos seus dados. Isto também leva a recomendação de palavras-chave mais adequadas, auxiliando na comunicação usuário/sistema.

4.6 VALIDAÇÃO DAS PALAVRAS-CHAVE POR PROFISSIONAIS

Os processos voltados para os dados de pesquisas, incluindo sua captura, preservação e disponibilização, surgem do fato que alguns dados gerados por estudos são únicos e dificilmente poderão ser reproduzidos ou reutilizados se forem perdidos ou avariados. Em algum ponto da história, a comunidade científica percebeu que apenas publicar os resultados de suas pesquisas através de artigos científicos e outros tipos de documentos não seria suficiente, sobretudo, quando os dados utilizados nessas pesquisas ficariam armazenados em arquivos pessoais e a mercê da limitação de vida útil das tecnologias recorrentes, sem a verdadeira intenção de preservá-los para a posteridade.

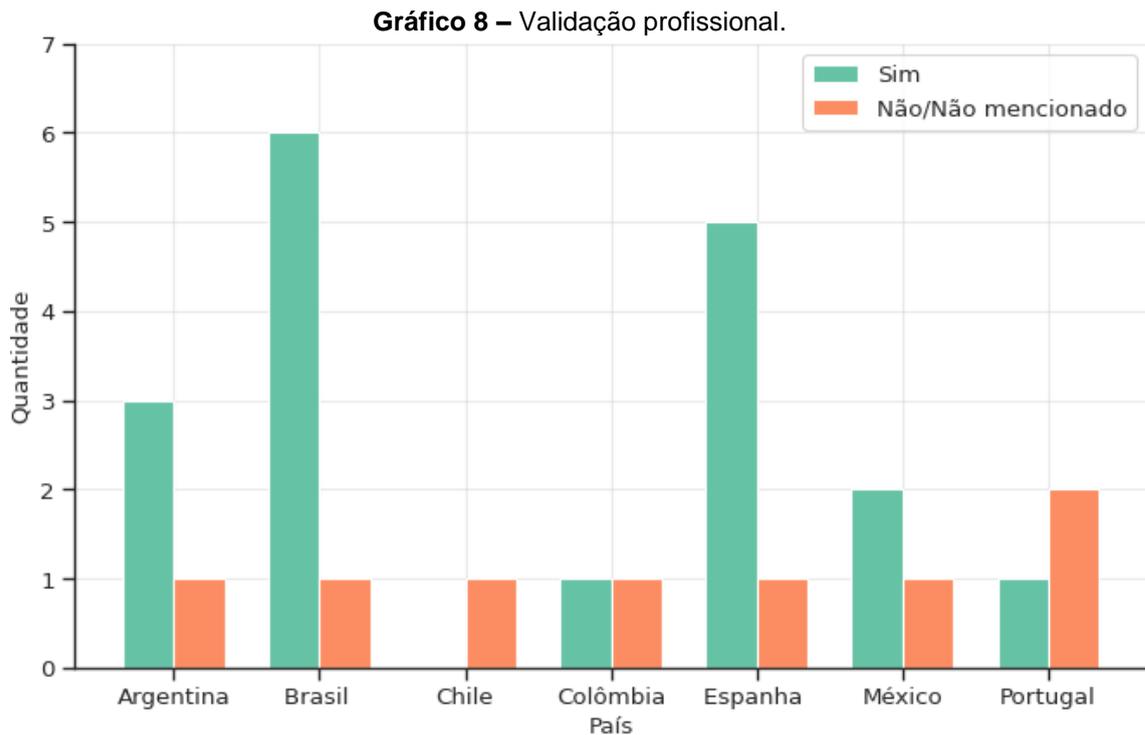
Nessa esfera, o surgimento de estratégias e metodologias aliadas a ambientes digitais que visassem à coleta, preservação e o compartilhamento de dados de pesquisa, fez-se necessário, pois esses ativos demandam cuidados característicos que envolvem “ações de arquivamento seguro, preservação, formas de acrescentar valor a esses conteúdos e de otimização da sua capacidade de reuso” (SAYÃO; SALES, 2012, p. 180).

A Curadoria Digital de Dados Científicos – CDDC tenciona cuidar dos dados científicos, assim, seu principal objetivo é “desenvolver estratégias para resolver problemas do fluxo da informação digital” (RESENDE; BAX, 2020, p. 95). Deve envolver ações dinâmicas que abarquem todos os processos, mesmo antes do depósito dos dados de pesquisa, auxiliando do planejamento até após o arquivamento, visto que objetiva manter os dados de pesquisa acessíveis em qualquer período de tempo.

Dessa maneira, levando em consideração seus objetivos, o curador é percebido como uma figura que zela pelos dados científicos, filtrando, organizando e protegendo, oferecendo diversas ações para que os dados de pesquisa se encontrem acessíveis.

Observa-se no Gráfico 8 que do total, 18 repositórios de dados de pesquisa identificados se preocupam em fazer a curadoria dos metadados. Destaca-se que essas diretrizes indicam especificamente a verificação e validação dos metadados por profissionais após o autoarquivamento, determinando que os dados de pesquisa arquivados sejam disponibilizados uma vez que os metadados se encontrem preenchidos de forma adequada.

Isto implica que a validação do metadado assunto preenchido pelo autor-pesquisador estão sendo realizadas, o que contribui para a recuperação da informação, pois a curadoria dos metadados eleva a qualidade dos mesmos, diminuindo o risco de termos inadequados que afetariam a encontrabilidade da informação.



Fonte: dados de pesquisa.

Por conseguinte, observa-se que

As infraestruturas de gestão de coleções de dados de pesquisa precisam estar ancoradas em uma política formal que defina todos os procedimentos do fluxo de curadoria: procedimentos práticos, formatos, metadados, arquivamento, preservação e segurança, licenças e, sobretudo, o tratamento que deve ser conferido ao compartilhamento de dados sensíveis, pessoais, confidenciais, de informações de interesse comercial ou que serão importantes nos processos de patenteamento, além das questões associadas à proteção de direitos de propriedade intelectual (SAYÃO; SALES, 2017).

Os dados de pesquisa que se encontram disponibilizados nesses ambientes estão estruturados por modelos de informação que se traduzem em esquemas de metadados. Assim, a curadoria que ocorre nesses repositórios será capaz de permitir que os dados de pesquisa estejam disponíveis para usuários futuros que

poderão reutilizar e reavaliar esses mesmos dados em novas circunstâncias, dando origem a novas pesquisas.

Portanto, a CDDC se constitui em um conjunto de ferramentas e iniciativas que aliadas às tecnologias é capaz de cuidar dos dados de pesquisa, o que leva não apenas os curadores estarem atentos à forma como os metadados estão sendo preenchidos, mas também o autor-pesquisador e os usuários futuros, pois os dados necessitam permanecerem interpretáveis ininterruptamente. Por consequência, “assegura a sustentabilidade dos dados para o futuro, não deixando, entretanto, de conferir valor imediato a eles para os seus criadores e para os seus usuários” (SAYÃO; SALES, 2017, p. 185).

5 CONSIDERAÇÕES FINAIS

A Ciência Aberta trouxe uma série de iniciativas que incentivaram e foram precursoras de ações que colocaram os dados de pesquisa em evidência, considerados como um dos produtos mais importantes provenientes de pesquisas científicas. Em meio a isso, várias ferramentas foram desenvolvidas para apoiar os processos envolvendo os dados de pesquisa, dentre as quais os repositórios de dados de pesquisa despontam como uma das principais.

Os repositórios de dados de pesquisa encontram seu propósito quando dão suporte à comunidade científica fornecendo meios para que os pesquisadores resguardem seus dados, ao passo que permitem o compartilhamento e fornecem visibilidade sobre os mesmos. À vista disso, esses ambientes digitais estão envolvidos em quase todo o ciclo de vida dos dados de pesquisa, sendo considerados fundamentais para a recuperação da informação, para o reuso de dados de pesquisa e para a reprodutibilidade científica.

A partir disso, incertezas a respeito da credibilidade desses ambientes se estabeleceram, o que levou a elaboração de parâmetros que fossem capazes de identificar aspectos relacionados às suas funcionalidades, como também que fossem capazes de considerar os usuários desses sistemas. Dessa maneira, foram instituídos os princípios *FAIR* como forma de assegurar que os metadados disponíveis nesses ambientes fossem identificados por computador e por usuário, promovendo a interoperabilidade entre diferentes sistemas que utilizam dados.

Os metadados abarcam todas as ações para as quais os princípios *FAIR* foram desenvolvidos, assim, eles necessitam ser localizáveis, acessíveis, interoperáveis e reutilizáveis a fim de que alcancem os níveis mais altos de *FAIRness* possíveis, considerando as particularidades de cada repositório de dados de pesquisa. Dentre eles, o metadado assunto permite que os conteúdos dos dados sejam identificados para os fins de representação e recuperação da informação pelo sistema, o que afetará diretamente a encontrabilidade da informação nesses ambientes digitais.

Para tanto, são atribuídas palavras-chave pelo autor-pesquisador no momento do depósito dos dados de pesquisa. As palavras-chave são capazes de traduzir os principais conceitos contidos nos dados de pesquisa, permitindo que o autor-pesquisador atue, também, como indexador dos seus próprios documentos.

Dessa forma, as palavras-chave são utilizadas para a representação da informação, possuindo um papel crucial na recuperação dos dados disponibilizados pelos repositórios de dados de pesquisa.

Posto isto, considerou-se pertinente investigar as diretrizes estabelecidas por repositórios de dados de pesquisa ibero-americanos com o intuito de identificar as diretrizes utilizadas para guiar o processo de atribuição de palavras-chave pelo autor-pesquisador na prática do autoarquivamento. No primeiro momento, foram apresentadas considerações acerca do movimento de acesso aberto e o autoarquivamento, repositórios de dados de pesquisa, dados de pesquisa, os princípios FAIR e sua aplicabilidade em repositórios de dados de pesquisa, assim como discussões envolvendo metadados, com especial atenção ao metadado assunto e a atribuição de palavras-chave para a representação de conteúdos em ambientes de repositórios digitais.

No segundo momento, a pesquisa se direcionou para a identificação dos repositórios de dados de pesquisa ibero-americanos que adotam o autoarquivamento e de suas respectivas diretrizes, mediante o uso da ferramenta Re3data. Mediante análises exaustivas dos documentos e dos ambientes digitais dos 26 repositórios de dados de pesquisa contemplados na pesquisa, observou-se a situação atual das diretrizes quanto ao preenchimento do metadado assunto.

Relativo à existência de diretrizes formalizadas, os repositórios de dados de pesquisa estão no caminho para a regularização de seus procedimentos. No entanto, o número ainda não pode ser considerado animador, visto que corresponde a apenas 61% dos repositórios pesquisados, levando em consideração o fato de nem todos oferecem instruções explícitas sobre o metadado assunto.

Ao analisar os recursos disponibilizados pelas páginas dos repositórios, constatou-se que, em alguns deles, os documentos contendo as diretrizes não se encontram facilmente dispostos para os usuários, sendo necessárias buscas exaustivas a fim de os encontrar, fato que pode desencorajar o autor-pesquisador a autoarquivar seus dados de pesquisa, da mesma forma que procurar compreender como é realizado esse processo.

A maioria dos repositórios utiliza o termo 'palavras-chave' para se referir ao campo dedicado à inclusão dos conteúdos dos dados de pesquisa. No entanto, observou-se que não é um campo padronizado, chegando até mesmo a não ser incluído de nenhuma forma em alguns repositórios de dados de pesquisa

contemplados nesta pesquisa. Isso se mostra preocupante, pois o metadado assunto é um dos principais metadados disponibilizados para a recuperação das informações dos dados de pesquisa, podendo refletir também em outras áreas, como quantidade de citações, entre outras.

Da mesma forma, a falta de instruções concretas para a atribuição de palavras-chave mostra-se um cenário preocupante, visto que delas dependem o preenchimento adequado do metadado assunto por parte do autor-pesquisador na prática do autoarquivamento, de forma que o sistema consiga recuperar documentos solicitados por usuários. Assim, mesmo que a maioria indique o uso de linguagens documentárias, esta orientação perde seu sentido ao não incluir em suas diretrizes instruções que orientem os autores-pesquisadores na condução da representação dos conteúdos dos dados de pesquisa.

Em todo caso, boa parte dos repositórios de dados de pesquisa aqui abordados realizam a curadoria dos metadados. Alguns até mesmo oferecem a opção de o arquivamento dos dados de pesquisa ser realizado pelo próprio repositório, se configurando com um método misto, no qual o arquivamento ser atribuído ao repositório fica a critério do autor-pesquisador. Porém, o arquivamento realizado pela instituição já se mostrou uma prática contraproducente, dado que este pode ser um processo dispendioso para os repositórios digitais, além de não colaborar com os preceitos estabelecidos pela Ciência Aberta.

De modo geral, os resultados são um tanto quanto negativos quando se trata das diretrizes estabelecidas por repositórios de dados de pesquisa ibero-americanos que adotam o autoarquivamento. Apesar de os preceitos da Ciência Aberta estarem em franca aceitação por parte da comunidade científica, colocá-las em prática mostra-se um pouco mais complexo. Por isso, estudos que busquem identificar as inconsistências e lacunas ainda vigentes no contexto dos repositórios digitais, sobretudo as que têm relação com a representação de conteúdos, são cada vez mais necessários.

Recomenda-se que aqueles repositórios de dados de pesquisa aqui identificados que possuem documentos já formalizados reavaliem suas diretrizes a fim de identificar se contemplam e oferecem orientações claras sobre o processo de atribuição de palavras-chave pelo autor-pesquisador. Para aqueles sistemas que não possuem esses documentos estabelecidos, que formalizem suas diretrizes com o propósito de oferecer orientações aos seus autores-pesquisadores.

Como proposta para trabalhos futuros, salienta-se que a investigação da qualidade do metadado assunto frente às diretrizes estabelecidas por repositórios de dados de pesquisa se mostra como um nicho interessante a ser explorado, uma vez que estudos nessa direção colaboram para identificar qual impacto as diretrizes contidas nos documentos disponibilizados realmente causam na representação de conteúdos de dados de pesquisa.

Discussões envolvendo o autoarquivamento em repositórios digitais é um tema em franco desenvolvimento, sendo oportuno que novas pesquisas se dediquem à questão. Ao descrever o processo de atribuição de palavras-chave para a representação do conteúdo de dados de pesquisa definido por cada um dos repositórios de dados de pesquisa analisados, os resultados da presente pesquisa contribuem para o desenvolvimento de estudos que ofereçam recomendações para a melhoria dos processos que envolvem o preenchimento do metadado assunto em sistemas de informação híbridos que adotam o autoarquivamento.

REFERÊNCIAS

- ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. NBR 6028. **Informação e documentação**: resumo. Rio de Janeiro: 2021. Disponível em: http://plone.ufpb.br/secretariado/contents/documentos/2021_ABNT6028Resumo.pdf. Acesso em: 26 nov. 2021.
- AMERICAN NATIONAL INSTITUTES FOR HEALTH. **NIH Strategic Plan for Data Science**. 2016. Disponível em: <https://commonfund.nih.gov/bd2k>. Acesso em: 27 fev. 2022.
- BETHESDA STATEMENT ON OPEN ACCESS PUBLISHING. 2003. Disponível em: https://www.researchgate.net/publication/48547523_The_Bethesda_Statement_on_Open-Access_Publishing. Acesso em: 20 nov. 2021.
- BJÖRK, B.-C. et al. Anatomy of green open access: Journal of the American Society for Information Science and Technology. **Journal of the Association for Information Science and Technology**, v. 65, n. 2, p. 237–250, fev. 2014. Disponível em: <https://asistdl.onlinelibrary.wiley.com/doi/abs/10.1002/asi.22963>. Acesso em: 24 nov. 2021.
- BUDAPEST OPEN ACCESS INITIATIVE. 2002. Disponível em: <https://www.budapestopenaccessinitiative.org/read/#>. Acesso em: 20 nov. 2021.
- BUCKLAND, M. K. Information as thing. **Journal of the American Society for Information Science**, v. 42, n. 5, p. 351–360, 1991. Disponível em: [https://asistdl.onlinelibrary.wiley.com/doi/abs/10.1002/\(SICI\)1097-4571\(199106\)42:5%3C351::AID-ASI5%3E3.0.CO;2-3](https://asistdl.onlinelibrary.wiley.com/doi/abs/10.1002/(SICI)1097-4571(199106)42:5%3C351::AID-ASI5%3E3.0.CO;2-3). Acesso em: 22 fev. 2021.
- CAPLAN, P. **Metadata fundamentals for all librarians**. Chicago: American Library Association, 2003. Disponível em: http://pustaka.unp.ac.id/file/abstrak_kki/EBOOKS/LIBRARIES%20Metadata%20fundamentals%20for%20all%20librarians.pdf. Acesso em: 07 nov. 2020.
- COSTA, S. M. S. Filosofia aberta, modelos de negócios e agências de fomento: elementos essenciais a uma discussão sobre o acesso aberto à informação científica. **Ciência da Informação**, v. 35, n. 2, p. 39–50, ago. 2006. Disponível em: <https://www.scielo.br/j/ci/a/XsgXnnC7xWHNR7gXrP9Hw3M/abstract/?lang=pt&format=html>. Acesso em: 25 nov. 2021.
- DATA FAIRPORT. **What is Data FAIRport initiative?** 2014. Disponível em: <https://www.datafairport.org/>. Acesso em: 15 mar. 2022.
- 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, 2007. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S0098791306001675>. Acesso em: 11 dez. 2020.

DUTCH TECHCENTRE FOR LIFE SCIENCES. **Jointly designing a Data FAIRPORT**. 2014. Disponível em: <https://www.dtls.nl/2014/01/20/jointly-designing-data-fairport/>. Acesso em: 15 mar. 2022.

EUROPEN COMISSION. **Commission High Level Expert Group on the European Open Science Cloud**. Realising the European open science cloud. Brussels: European Commission, 2016. Disponível em: https://ec.europa.eu/research/openscience/pdf/realising_the_european_open_science_cloud_2016.pdf. Acesso em: 15 mar. 2022.

EUROPEN COMISSION. **Guidelines on Open Accessto Scientific Publications and Research Datain Horizon 2020**. V2.1. 2020. Disponível em: <https://www.fosteropenscience.eu/content/guidelines-open-access-scientific-publications-and-research-data-horizon-2020>. Acesso em: 15 mar. 2022.

FORMENTON, D. et al. Os patronos dos metadados como recursos tecnológicos para a garantia da preservação digital. **Biblios**, Pittsburgh, n. 68, pág. 82-95, 2017. Disponível em: http://www.scielo.org.pe/scielo.php?script=sci_arttext&pid=S1562-47302017000300006&lng=es&nrm=iso. Acesso em: 20 dez. 2020.

FOULONNEAU, M.; RILEY, J. **Metadata for digital resources: implementation, systems design and interoperability**. Oxford: Chandos Publishing, 2008.

FREITAS, M. P. **Autoarquivamento e representação de assunto: estudo analítico de teses e dissertações do Repositório Institucional da UFSCar**. 2019. Dissertação (mestrado) - Universidade Federal de São Carlos, Centro de Educação e Ciências Humanas, Programa de Pós-graduação em Ciência da Informação, São Carlos, 2019. Disponível em: http://btdtd.ibict.br/vufind/Record/SCAR_2bc79be369790de4267a60073367925d. Acesso em: 5 jan. 2021.

FREITAS, M. P.; DAL'EVEDOVE, P. R.; TARTAROTTI, R. C. Políticas de autoarquivamento em repositórios institucionais brasileiros: estudo analítico do metadado assunto. **Páginas a&b**, 3, 169-175. 2021. Disponível em: <http://aleph20.letras.up.pt/index.php/paginasaeb/article/view/10245>. Acesso em: 15 jan. 2021.

FUJITA, M. S. L.; TARTAROTTI, R. C. D. Análise de palavras-chave da produção científica de pesquisadores: o autor como indexador. **Informação & Informação**, v. 25, n. 3, p. 332-374, 2020. Disponível em: <https://brapci.inf.br/index.php/res/v/149488>. Acesso em: 26 nov. 2021.

GAMA, I. S. et al. Tendências e perspectivas de pesquisa sobre repositórios digitais no Brasil: uma análise de Rede Sociais (ARS). **Revista Eletrônica Comunicação, Informação e Inovação em Saúde**, v. 11, p. 1–14, 2017. Disponível em: <https://www.arca.fiocruz.br/handle/icict/23734>. Acesso em: 07 nov. 2020.

GONÇALVES, A. L. Uso de resumos e palavras-chave em Ciências Sociais: uma avaliação. **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação**, [S. l.], v. 13, n. 26, p. 78-93, 2008. Disponível em:

<https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2008v13n26p78>. Acesso em: 26 nov. 2021.

GREENBERG, J. Understanding Metadata and Metadata Schemes. **Cataloging & Classification Quarterly**, v. 40:3-4, p. 17-36, 2005. Disponível em: https://www.tandfonline.com/doi/abs/10.1300/J104v40n03_02. Acesso em: 13 nov. 2020.

HARNAD, S. **Scholarly journals at the crossroads: a subversive proposal for electronic publishing**. Washington: Association of Research Libraries, 1995. Disponível em: <https://eprints.soton.ac.uk/362894/1/subversive-proposal-electronic-publishing-jun05.pdf>. Acesso em: 5 dez. 2021.

HARTLEY, J.; KOSTOFF, R. N. How useful are 'keywords' in scientific journals? **Journal of information science**, v. 29, n. 5, p. 433-438, 2003. Disponível em: <https://journals.sagepub.com/doi/10.1177/01655515030295008>. Acesso em: 24 nov. 2021.

HENNING, P. C. et. al. 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, 2019a. Disponível em: <https://brapci.inf.br/index.php/res/v/113770>. Acesso em: 10 mar. 2022.

HENNING, P. C. et. al. Desmistificando os princípios fair: conceitos, métricas, tecnologias e aplicações inseridas no ecossistema dos dados fair. **Encontro Nacional de Pesquisa e Pós-graduação em Ciência da Informação**, n. XIX ENANCIB, 2019b. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/103243>. Acesso em: 10 mar. 2022.

HERMANN, G. **Padrões de Metadados**. Infraestrutura Nacional de Dados Abertos (INDA), 2017. Disponível em: <https://wiki.dados.gov.br/Padroes-de-metadados.ashx#:~:text=O%20padr%C3%A3o%20de%20metadados%20da,para%20os%20utilizadores%20dos%20dados>. Acesso em: 13 nov. 2020.

IKEMATU, R. S. Gestão de metadados: sua evolução na tecnologia da informação. **DataGramZero**, v. 2, n. 6, 2001. Disponível em: https://brapci.inf.br/_repositorio/2010/01/pdf_0a6da12dc0_0007454.pdf. Acesso em: 10 jan. 2021.

IMMING, M. **FAIR Data Advanced Use Cases: from principles to practice in the Netherlands (Final)**. 2018. Disponível em: <https://zenodo.org/record/1250535#.Wyl3RkiFNaQ>. Acesso em: 19 mar. 2022.

INSTITUTO FEDERAL GOIANO. **Repositório de Dados Brutos**. 2020. Disponível em: <https://www.ifgoiano.edu.br/home/index.php/publicacoes-pesquisa-urutai/15627-repositorio-de-dados-brutos.html>. Acesso em: 13 out. 2021.

JOHNSTON, L. R. Introduction to Data Curation. *In*: JOHNSTON, L. R. (ed). **Curating Research Data: Practical Strategies for Your Digital Repository**. Chicago: Association of College and Research Libraries, 2017, v.1. p. 1-30. Disponível em:

https://www.ala.org/acrl/sites/ala.org.acrl/files/content/publications/booksanddigitalresources/digital/9780838988596_crd_v1_OA.pdf. Acesso em: 22 fev. 2021.

LARA, M. L. G. Linguagem documentária e terminologia. **Transinformação**, v. 16, n. 3, p. 231–240, dez. 2004. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0103-37862004000300003&lng=pt&tlng=pt. Acesso em: 16 mar. 2022.

LYNCH, C. A. Institutional repositories: essential infrastructure for scholarship in the digital age. **Portal: Libraries and the Academy**, v. 3, n. 2, p. 327-336, 2003. Disponível em: <https://muse.jhu.edu/article/42865>. Acesso em: 23 mar. 2022.

MEDEIROS, G. M. de. **Organização da informação em repositórios digitais: implicações do autoarquivamento na representação da informação**. 2010. Dissertação (mestrado) - Universidade Federal de Santa Catarina, Centro de Ciências da Educação, Programa de Pós-graduação em Ciência da Informação, Florianópolis, 2010. Disponível em: <https://repositorio.ufsc.br/handle/123456789/94615>. Acesso em: 15 jan. 2021.

MIGUÉIS, A. et al. A importância das palavras-chave dos artigos científicos da área das Ciências Farmacêuticas, depositados no Estudo Geral: estudo comparativo com os termos atribuídos na MEDLINE. **InCID: Revista de Ciência da Informação e Documentação**, [S. l.], v. 4, n. 2, p. 112-125, 2013. DOI: 10.11606/issn.2178-2075.v4i2p112-125. Disponível em: <https://www.revistas.usp.br/incid/article/view/69284>. Acesso em: 20 nov. 2021.

NATIONAL RESEARCH COUNCIL. **A Question of Balance: Private Rights and the Public Interest in Scientific and Technical Databases**. Washington, D.C.: National Academies Press, 1999. Disponível em: <https://www.nap.edu/catalog/9692/a-question-of-balance-private-rights-and-the-public-interest>. Acesso em: 3 mar. 2021.

NORUZI, A. Folksonomies: Why do we need controlled vocabulary? **Webology**, v. 4, n. 2, 2007. Disponível em: <http://eprints.rclis.org/10308/>. Acesso em: 20 mar. 2022.

OLIVEIRA, L. P. Política de Indexação: Concepções acerca do conceito e percepções em torno de sua elaboração. **Ciência da Informação em Revista**, v. 4, n. 2, 2017. Disponível em: <https://www.seer.ufal.br/index.php/cir/article/view/3463>. Acesso em: 19 jul. 2020.

OLIVEIRA, L. P. et al. Política de indexação em periódicos da Ciência da Informação: um estudo das diretrizes para atribuição de palavras-chave aos artigos. **Perspectivas em Ciência da Informação**, v. 25, p. 140-169, 2020. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1413-99362020000400140&tlng=pt. Acesso em: 20 nov. 2021.

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT. **Principles and Guidelines for Access to Research Data from Public Funding**. Paris, France: OECD, 2007. Disponível em: https://read.oecd-ilibrary.org/science-and-technology/oecd-principles-and-guidelines-for-access-to-research-data-from-public-funding_9789264034020-en-fr#page1. Acesso em: 3 mar. 2021.

PIOVEZAN, L. B. **Avaliação da indexação em catálogos de bibliotecas universitárias por meio da recuperação da informação**. 2015. Dissertação (Mestrado em Ciência da Informação) - Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2015. Disponível em: <http://repositorio.unesp.br/handle/11449/126586>. Acesso em: 07 nov. 2020.

RESENDE, L. C. DE; BAX, M. P. A curadoria de dados científicos na Ciência da Informação: levantamento do cenário nacional. **AtoZ: novas práticas em informação e conhecimento**, v. 9, n. 1, p. 94, 16 set. 2020. Disponível em: <https://revistas.ufpr.br/atoz/article/view/69190>. Acesso em: 17 mar. 2022.

RILEY, J. **Seeing Standards: A Visualization of the Metadata Universe**. 2009. Disponível em: <http://jennriley.com/metadatamap/>. Acesso em: 23 nov. 2020

RODRIGUES, E. et al. **Os repositórios de dados científicos: estado da arte**. 2010. Disponível em: <https://repositorio-aberto.up.pt/bitstream/10216/23806/2/44632.pdf>. Acesso em: 7 nov. 2020

RODRIGUES, M. M. **Repositórios de dados científicos na América do Sul: uma análise da conformidade com os princípios FAIR**. 2020. 111 f. Dissertação (Mestrado em Gestão e Organização do Conhecimento) – Escola de Ciência da Informação, Universidade Federal de Minas Gerais. Belo Horizonte, 2020. Disponível em: <https://repositorio.ufmg.br/handle/1843/34774>. Acesso em: 19 mar. 2022.

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) - Escola de Comunicação, Universidade Federal do Rio de Janeiro, 2014. Disponível em: <https://ridi.ibict.br/handle/123456789/874>. Acesso em: 8 nov. 2020.

SALES, L. F.; SAYÃO, L. F. Uma proposta de taxonomia para dados de pesquisa. **Revista Conhecimento em Ação**, v. 4, n. 1, p. 31–48, 30 jun. 2019. Disponível em: <https://revistas.ufrj.br/index.php/rca/article/view/26337>. Acesso em: 22 jun. 2021.

SANTOS, F. C. L. **Coerência na representação temática de artigos científicos indexados no repositório de saúde pública da Fundação Oswaldo Cruz**. 2017a. 258 f. Dissertação (Mestrado em Ciência da Informação) – Universidade Federal do Rio de Janeiro/Escola de Comunicação, Instituto Brasileiro de Informação em Ciência e Tecnologia, Programa de Pós-Graduação em Ciência da Informação, Rio de Janeiro, 2017. Disponível em: <https://www.arca.fiocruz.br/handle/icict/25837>. Acesso em: 7 nov. 2020.

SANTOS, R. F.; NEVES, D. A. B. Práticas de indexação em repositórios digitais de acesso aberto: análise do metadado assunto do repositório institucional da UFRN. **XIX Encontro Nacional de Pesquisa em Ciência da Informação (XIX ENANCIB)**, v. 24, n. 2, 2018. Disponível em: <https://brapci.inf.br/index.php/res/v/102563>. Acesso em: 7 nov. 2020.

SAYÃO, L. F.; SALES, L. F. Curadoria digital: um novo patamar para preservação de dados digitais de pesquisa. **Informação & Sociedade: Estudos**, 2012. Disponível

em: <https://brapci.inf.br/index.php/res/v/92680>. Acesso em: 20 mar. 2022.

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. Disponível em:

<https://brapci.inf.br/index.php/article/view/0000022200/59fad3d050cb444b44e1a272ce1e0ca3>. Acesso em: 8 nov. 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, 9 jan. 2017. Disponível em: <http://revistas.ufpr.br/atoz/article/view/49708>. Acesso em: 20 mar. 2022.

SAYÃO, L. F.; SALES, L. F. Afinal, o que é Dado de Pesquisa? **Biblos**, [S. l.], v. 34, n. 2, 2020. Disponível em: <https://periodicos.furg.br/biblos/article/view/11875>. Acesso em: 26 jun. 2021.

SCIENTIFIC ELECTRONIC LIBRARY ONLINE. **Guia de preparação de dados de pesquisa**. 2021. Disponível em: https://wp.scielo.org/wp-content/uploads/Guia_preparacao_pt.pdf. Acesso em: 19 mar. 2022.

SEMLER, 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, 3 maio 2019. Disponível em: <http://revista.ibict.br/ciinf/article/view/4461>. Acesso em: 13 jun. 2020.

SILVA, F. C. C.; SILVEIRA, L. O ecossistema da Ciência Aberta. **Transinformação**, v. 31, p. e190001, 2019. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0103-37862019000100302&tlng=pt. Acesso em: 22 mar. 2022.

SMIRAGLIA, R. P. Introducing metadata. **Cataloging & Classification Quarterly**, v. 40, n. 3-4, p. 1-15, 2005. Disponível em: https://www.tandfonline.com/doi/abs/10.1300/J104v40n03_01?journalCode=wccq2. Acesso em: 7 nov. 2020.

SOCIEDADE MAX PLANCK. **Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities**. 2003. Disponível em: <https://openaccess.mpg.de/Berlin-Declaration>. Acesso em: 20 nov. 2021.

SOCIEDADE MAX PLANCK. **Progress in Implementing the Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities**. 2005. Disponível em: https://openaccess.mpg.de/319838/Berlin_3. Acesso em: 20 nov. 2021.

SWAN, A. Why Open Access for Brazil? Por que acesso livre no Brazil? **Liinc em Revista**, [S. l.], v. 4, n. 2, 2008. Disponível em: <http://revista.ibict.br/liinc/article/view/3173>. Acesso em: 26 nov. 2021.

SWANSON, J.; RINEHART, A. K. Data in context: Using case studies to generate a common understanding of data in academic libraries. **The Journal of Academic Librarianship**, v. 42, n. 1, p. 97–101, jan. 2016. Disponível em:

<https://kb.osu.edu/handle/1811/82202>. Acesso em: 13 jun. 2021.

TARTAROTTI, R. C. D. **Avaliação do processo de indexação de assuntos em repositórios institucionais pela abordagem da recuperação da informação**. 2019. 370 f. Tese (Doutorado em Ciência da Informação) Faculdade de Filosofia e Ciências, Universidade Estadual Paulista Júlio de Mesquita Filho, Programa de Pós-Graduação em Ciência da Informação, Marília, 2019. Disponível em: <https://repositorio.unesp.br/handle/11449/191064>. Acesso em: 7 nov. 2020.

TARTAROTTI, R.C.D.; DAL'EVEDOVE, P.R.; FUJITA, M.S.L. Indexação de assuntos em repositórios digitais: uma incursão na produção científica brasileira de Ciência da Informação. **Scire**: representação e organização do conhecimento, [S. l.], v. 26. 2 P. 13–18, 2020. DOI: 10.54886/scire.v26i2.4626. Disponível em: <https://ojs.ibersid.eu/index.php/scire/article/view/4626>. Acesso em: 20 nov. 2021.

TERRA, A. L. et al. Subject-access metadata on ETD supplied by authors: A case study about keywords, titles and abstracts in a Brazilian academic repository. **Journal of Academic Librarianship**, v. 47, p. 102268, 2021. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S0099133320301592>. Acesso em: 25 nov. 2021.

VEIGA, V.; MACENA, L. G. O autoarquivamento nos Repositórios institucionais brasileiros: um estudo exploratório. **Ponto de Acesso**, v. 9, n. 3, p. 35-47, 2015. Disponível em: <http://dx.doi.org/10.9771/1981-6766rpa.v9i3.15107>. Acesso em: 20 nov. 2021.

VIANA, J. M. A.; DAL'EVEDOVE, P. R. Indexação de Dados Científicos: uma análise a partir das políticas da Rede de Repositórios de Dados Científicos do Estado de São Paulo. **Páginas a&mp**: arquivos e bibliotecas, [S. l.], p. 192–196, 2021. Disponível em: <https://ojs.letras.up.pt/index.php/paginasueb/article/view/10041>. Acesso em: 20 nov. 2021.

VILLALOBOS, A. P. O.; GOMES, F. A. Padronização de metadados na representação da informação em repositórios institucionais de universidades federais brasileiras. In: **XVI Encontro Nacional de Pesquisa em Pós-Graduação em Ciência da Informação**. 2015. Disponível em: <http://www.ufpb.br/evento/index.php/enancib2015/enancib2015/paper/view/264/0>. Acesso em: 7 nov. 2020.

WILKINSON, M. D. et al. The FAIR Guiding Principles for scientific data management and stewardship. **Scientific data**, v. 3, n. 1, p. 1-9, 2016. Disponível em: <https://www.nature.com/articles/sdata201618>. Acesso em: 10 mar. 2022.

ANEXO A – Identificação dos padrões de metadados por repositório de dados de pesquisa

País	Repositório de dados de pesquisa	Padrão de Metadado utilizado
Argentina	Produção acadêmica UCC	Dublin Core
Argentina	Portal de datos de Biodiversidad	Darwin Core
Argentina	SEDICI	Dublin Core
Brasil	Humadoc repositório	Dublin Core
Brasil	Globe	ABCD
Brasil	PPBio Data Repository	EML
Brasil	IBICT Cariniana Dataverse	DDI
Brasil	SciELOData	DDI
Brasil	SciELOData	DataCite Metadata Schema
Brasil	CEDAP Research data repository	Dublin Core
Brasil	Repositório Institucional UNESP	Indefinido
Brasil	Repositório de dados de pesquisa do Instituto Federal Goiano - campus Urutaí	DataCite Metadata Schema
Brasil	Repositório de dados de pesquisa do Instituto Federal Goiano - campus Urutaí	Dublin Core
Brasil	Repositório de dados de pesquisa do Instituto Federal Goiano - campus Urutaí	DDI
Chile	Repositório de dados de pesquisa do Instituto Federal Goiano - campus Urutaí	OAI-ORE
Chile	Repositorio de datos de investigación de la Universidad de Chile	DDI
Chile	Repositorio de datos de investigación de la Universidad de Chile	ABCD
Colômbia	Repositorio de datos de investigación de la Universidad de Chile	Dublin Core
Colômbia	CIAT Dataverse	DDI
Colômbia	Repositorio de datos de investigación de la universidad del Rosario	Dublin Core
Colômbia	Repositorio de datos de investigación de la universidad del Rosario	DDI
Espanha	Repositorio de datos de investigación de la universidad del Rosario	DataCite Metadata Schema
Espanha	IoChem – BD	Dublin Core
Espanha	UPCommons - Research Data	DataCite Metadata Schema
Espanha	UPCommons - Research Data	Dublin Core
Espanha	O2 Dades	Dublin Core
Espanha	Digibug	Dublin Core
Espanha	RiuNet	OAI-ORE
Espanha	RiuNet	Dublin Core
Espanha	RiuNet	DataCite Metadata Schema
Espanha	Repositório Digital de Documentos UAB	Dublin Core
México	Repositório Digital de Documentos UAB	DataCite Metadata Schema
México	CIMMYT	DDI
México	ciRcle	Dublin Core
Portugal	RU-IIEc	Dublin Core
Portugal	DataRepositorium	OAI-ORE
Portugal	DataRepositorium	DDI
Portugal	DataRepositorium	Dublin Core
Portugal	DataRepositorium	DataCite Metadata Schema
Portugal	Kinetic models of biological systems	DDI
Portugal	Kinetic models of biological systems	Dublin Core

Portugal	DadosIPB	Dublin Core
Portugal	DadosIPB	DDI
Portugal	DadosIPB	DataCite Metadata Schema

ANEXO B – Diretrizes e demais documentos utilizados para análise

País	Repositório de dados de pesquisa	Documento/Fonte
	Produção acadêmica UCC ²⁰	Política del Repositorio Institucional - Producción Académica
Argentina	Portal de datos de Biodiversidad ²¹	Tutorial para completar los campos de METADATOS en IPT
	SEDICI ²²	Tutorial Para Realizar Autoarchivo
	Humadoc repositorio	Página do repositório/Dspace
	Globe ²³	Globe Batch Import Template
Brasil	PPBio Data Repository	Página do repositório
	IBICT Cariniana Dataverse ²⁴	Guia para os Usuários do Repositório Dataverse do IBICT/Dataverse
	ScieloData ²⁵	Guia de depósito de dados de pesquisa
	CEDAP Reasearch data repository	Página do repositório/Dspace
	Repositório Institucional UNESP ²⁶	Manual de Autoarquivamento de Dados de Pesquisa
	Repositório de dados de pesquisa do Instituto Federal Goiano - campus Urutaí	Página do repositório/Dataverse
Chile	Repositorio de datos de investigación de la Universidad de Chile	Página do repositório/Dataverse
Colômbia	CIAT Dataverse	Página do repositório/Dataverse
	Repositorio de datos de investigacion de la universidade del Rosario ²⁷	Política Institucional para la gestión de datos de investigación/Dataverse
Espanha	IoChem – BD	Página do repositório/Dspace
	UPCommons - Research Data	Página do repositório/Dspace
	O2 Dades ²⁸	Guia per a dipositar recerca a l'O2 repositori UOC
	Digibug ²⁹	Guía de Autoarchivo
	RiuNet ³⁰	Envíos a RiuNet
	Repositório Digital de Documentos UAB ³¹	Com lliurar documents al Dipòsit Digital de Documents de la UAB

²⁰ Produção acadêmica UCC. Disponível em: <http://pa.bibdigital.ucc.edu.ar/policies.html>.

²¹ Portal de datos de Biodiversidad. Disponível em: <https://www.gbif.es/wp-content/uploads/2018/05/Como-completar-los-metadatos-en-IPT-v.2.pdf>.

²² SEDICI. Disponível em: http://sedici.unlp.edu.ar/bitstream/handle/10915/26623/Documento_completo__versi%C3%B3n_actualizada_.pdf?sequence=5&isAllowed=y.

²³ Globe. Disponível em: <http://globe.umbc.edu/documentation-overview/representativeness/>.

²⁴ IBICT Cariniana Dataverse. Disponível em: https://cariniana.ibict.br/images/artigos/Dataverse/Guia_Usuarios_Dataverse_Ibict.pdf.

²⁵ ScieloData. Disponível em: https://wp.scielo.org/wp-content/uploads/Guia_deposito_pt.pdf.

²⁶ Repositório Institucional UNESP. Disponível em: <https://repositorio.unesp.br/>.

²⁷ Repositorio de datos de investigación de la Universidad de Chile. Disponível em: <https://repository.urosario.edu.co/handle/10336/19340>.

²⁸ O2 Dades. Disponível em: http://openaccess.uoc.edu/webapps/o2/bitstream/10609/21381/6/Guia%20GIR_O2-cat_CA.pdf.

²⁹ Digibug. Disponível em: <https://digibug.ugr.es/page/pasosAutoarchivo>.

³⁰ RiuNet. Disponível em: <https://riunet.upv.es/bitstream/handle/10251/17469/guia-env%c3%ados-a-RiuNet1.pdf?sequence=1&isAllowed=y>.

³¹ Repositório Digital de Documentos UAB. Disponível em: https://ddd.uab.cat/pub/guibib/136736/autoarxiudd_a2021m7.pdf.

México	CIMMYT ³²	Research Data and Information Product Management
	cIRcle ³³	cIRcle Metadata Manual/Dataverse
	RU-IIEc ³⁴	Manual para el Depósito de Objetos en el Repositorio Universitario del Instituto de Investigaciones Económicas
Portugal	DataRepositorium	Página do repositório/Dataverse
	Kinetic models of biological systems ³⁵	User Guide - <i>Version 2.0</i>
	DadosIPB	Página do repositório/Dataverse

³² CIMMYT. Disponível em: <https://www.cimmyt.org/content/uploads/2019/10/CIMMYT-Research-Data-and-Information-Product-Management-policy-2019-10.pdf>.

³³ cIRcle. Disponível em: https://wiki.ubc.ca/images/9/9a/CIRcle_Metadata_Manual_2022-02-03.pdf.

³⁴ RU-IIEc. Disponível em: https://ru.iiec.unam.mx/91/4/manual_deposito_documentos_ru.pdf.

³⁵ Kimosys. Disponível em: https://kimosys.org/documentation/Documentation_Kimosys_v2.pdf.