



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

---

DISSERTAÇÃO DE MESTRADO

**INDEXAÇÃO DE DADOS DE PESQUISA  
PELO AUTOR CIENTÍFICO:  
análise de diretrizes de sistemas de autoarquivamento  
do Brasil, Espanha e Portugal**

CESAR AUGUSTO LOPES

---

SÃO CARLOS  
2024

CESAR AUGUSTO LOPES

---

**INDEXAÇÃO DE DADOS DE PESQUISA  
PELO AUTOR CIENTÍFICO:  
análise de diretrizes de sistemas de autoarquivamento  
do Brasil, Espanha e Portugal**

---

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

**Financiamento:** Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES)

SÃO CARLOS  
2024

César Augusto, Lopes

Indexação de dados de pesquisa pelo autor científico:  
análise de diretrizes de sistemas de autoarquivamento  
do Brasil, Espanha e Portugal / Lopes César Augusto --  
2024.  
123f.

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, Januário Albino Nhacuongue  
Bibliografia

1. Indexação. 2. Dados de pesquisa . 3. Sistemas de  
autoarquivamento . I. César Augusto, Lopes. II. Título.

Ficha catalográfica desenvolvida pela Secretaria Geral de Informática  
(SIn)

DADOS FORNECIDOS PELO AUTOR

Bibliotecário responsável: Arildo Martins - CRB/8 7180



**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 do candidato Cesar Augusto Lopes, realizada em 29/07/2024.

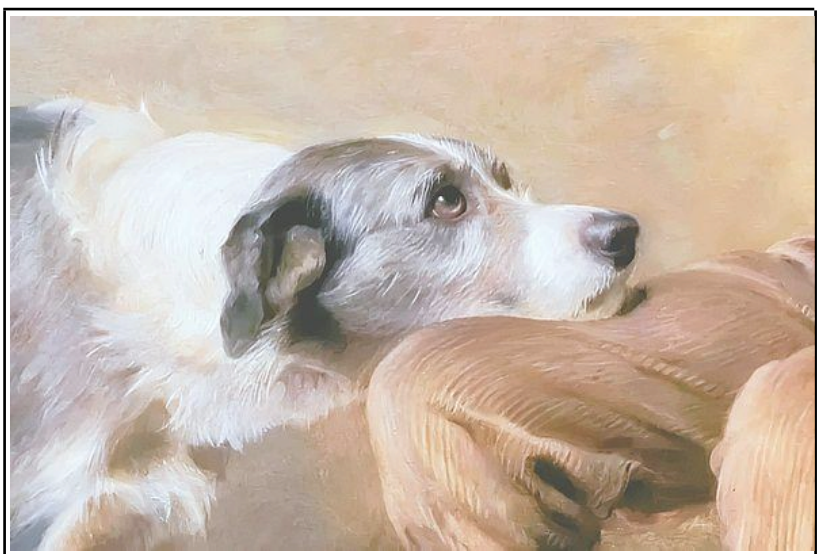
**Comissão Julgadora:**

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

Prof. Dr. Januario Albino Nhacuongue (UFSCar)

Profa. Dra. Ana Lúcia Silva Terra (UC)

O Relatório de Defesa assinado pelos membros da Comissão Julgadora encontra-se arquivado junto ao Programa de Pós-Graduação em Ciência da Informação.



Recorte de "*Fidelity*" (Briton Rivière, 1869)

À **Belinha**.

## AGRADECIMENTOS

À **minha orientadora, Profa. Dra. Paula Regina Dal'Evedove**, que mais uma vez, assim como durante a minha Iniciação Científica na Graduação, não se intimidou a trilhar comigo esta tempestuosa trajetória, rumo ao alcance da importante realização acadêmica e pessoal que hoje se consolida; alguém que, como um dia dito, escolheu “*segurar minha a mão pelo tempo que eu quisesse segurar a sua*”, superando eventuais limitações e dificuldades, sem abnegar da recíproca compreensão e paciência;

Aos *membros da Banca Examinadora* pelas ricas sugestões que refinaram o conteúdo desta Dissertação, conferindo uma maior qualidade ao final: **Profa. Dra. Ana Lúcia Terra** e **Prof. Dr. Januário Albino Nhacuongue**;

À **Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES)** pelo financiamento desta pesquisa; contributo que, admito, foi o combustível e o estímulo para um estudante desmotivado, por algumas vezes;

À **minha psicoterapeuta, Dra. Julisa Morales Calve**, por ter sido minha guia em vários trechos desse caminho e pela sua abundante generosidade no decorrer da minha formação;

À *mim mesmo*.

*“Fiz o que pude nas condições que eu tinha”.  
(autoria desconhecida)*

*“Sorte é o que acontece quando a preparação encontra a oportunidade”.  
(Sêneca)*

*“Aqueles que fogem do inferno, nunca falam sobre isso. E, depois de terem escapado,  
não há mais nada que os incomode”.  
(Charles Bukowski)*

## RESUMO

O acesso irrestrito aos dados de pesquisa em repositórios digitais representa um dos pilares da Ciência Aberta. Nesse contexto, a atribuição adequada de metadados temáticos durante o arquivamento desses dados é fundamental, pois facilita a sua recuperação e reutilização. Para garantir a eficácia desse processo, os repositórios digitais, especialmente aqueles que permitem práticas de autoarquivamento, devem fornecer aos autores científicos documentos com diretrizes claras para a atribuição de termos descritores e palavras-chave aos dados depositados. Partindo dessa premissa, surge a seguinte questão: *Como as diretrizes de repositórios de dados de pesquisa vinculados ao Brasil, Espanha e Portugal orientam os autores científicos na atribuição de palavras-chave durante o processo de autoarquivamento?* O objetivo principal deste estudo é *analisar os critérios orientadores presentes nas diretrizes para a indexação de dados de pesquisa em repositórios digitais vinculados a essas três nações, buscando, com isso, contribuir para a formalização de procedimentos de indexação mais estruturados e transparentes.* A pesquisa adota uma abordagem aplicada, qualitativa, descritiva e exploratória, com coleta de dados bibliográficos e documentais. O universo da pesquisa foi composto por repositórios digitais identificados no diretório re3data. A análise de documentos provenientes de 21 (vinte e um) repositórios qualificados para investigação revelou que nenhum desses sistemas de autoarquivamento apresenta um documento formal, especificamente intitulado "Política de Indexação", que trate da representação do conteúdo temático dos dados de pesquisa. Embora tal documento não tenha sido disponibilizado, ele poderia servir como base para o desenvolvimento de materiais complementares, como Manuais de Indexação, Guias de Autoarquivamento e Instruções para a Atribuição de Palavras-chave. Com o intuito de cumprir o objetivo da pesquisa, foram identificados e analisados 15 (quinze) critérios orientadores para a atribuição de palavras-chave, definidos com base no referencial teórico da pesquisa e nas observações dos documentos analisados. Constatou-se que as diretrizes dos repositórios examinados não abordam de forma completa todos os aspectos essenciais para a atribuição eficaz de palavras-chave pelos autores científicos. As instruções fornecidas são frequentemente superficiais e carecem de profundidade e relevância, com documentos que não apresentam procedimentos suficientemente estruturados e claros. A conclusão do estudo aponta que a melhoria da qualidade dos documentos orientadores para a indexação, especialmente em práticas de autoarquivamento, deve ser uma prioridade para as instituições responsáveis pela gestão desses repositórios. Os critérios apresentados nesta pesquisa podem servir como base para a criação de documentos mais eficientes e robustos, capazes de orientar de maneira mais precisa os autores científicos na representação temática dos dados de suas pesquisas.

**Palavras-chave:** Indexação. Dados de pesquisa. Palavras-chave. Autoarquivamento. Repositórios digitais.

## ABSTRACT

Unrestricted access to research data in digital repositories represents one of the pillars of Open Science. In this context, the proper assignment of thematic metadata during the archiving of such data is crucial, as it facilitates their retrieval and reuse. To ensure the effectiveness of this process, digital repositories, especially those that allow self-archiving practices, must provide scientific authors with documents containing clear guidelines for the assignment of descriptor terms and keywords to deposited data. Based on this premise, the following question arises: *How do the guidelines of research data repositories linked to Brazil, Spain, and Portugal guide scientific authors in assigning keywords during the self-archiving process?* The main objective of this study is to analyze the guiding criteria present in the guidelines for the indexing of research data in digital repositories associated with these three nations, aiming to contribute to the formalization of more structured and transparent indexing procedures. The research adopts an applied, qualitative, descriptive, and exploratory approach, with bibliographic and documental data collection. The research universe consisted of digital repositories identified in the re3data directory. The analysis of documents from 21 (twenty-one) repositories qualified for investigation revealed that none of these self-archiving systems provides a formal document specifically titled "Indexing Policy" that addresses the thematic content representation of research data. Although such a document was not made available, it could serve as a basis for the development of complementary materials, such as Indexing Manuals, Self-Archiving Guides, and Instructions for Assigning Keywords. To meet the research objective, 15 (fifteen) guiding criteria for the assignment of keywords were identified and analyzed, defined based on the theoretical framework of the study and observations from the analyzed documents. It was found that the guidelines of the examined repositories do not comprehensively address all the essential aspects of effective keyword assignment by scientific authors. The instructions provided are often superficial and lack depth and relevance, with documents that do not present sufficiently structured and clear procedures. The study concludes that improving the quality of guiding documents for indexing, especially in self-archiving practices, should be a priority for the institutions responsible for managing these repositories. The criteria presented in this research can serve as a foundation for creating more efficient and robust documents capable of more accurately guiding scientific authors in the thematic representation of their research data.

**Keywords:** Indexing. Research data. Keywords. Self-archiving. Digital repositories.

## RESUMEN

El acceso irrestricto a los datos de investigación en repositorios digitales representa uno de los pilares de la Ciencia Abierta. En este contexto, la asignación adecuada de metadatos temáticos durante el archivado de estos datos es fundamental, ya que facilita su recuperación y reutilización. Para garantizar la eficacia de este proceso, los repositorios digitales, especialmente aquellos que permiten prácticas de autoarchivo, deben proporcionar a los autores científicos documentos con directrices claras para la asignación de términos descriptivos y palabras clave a los datos depositados. Partiendo de esta premisa, surge la siguiente pregunta: *¿Cómo orientan las directrices de los repositorios de datos de investigación vinculados a Brasil, España y Portugal a los autores científicos en la asignación de palabras clave durante el proceso de autoarchivo?* El objetivo principal de este estudio es analizar los criterios orientadores presentes en las directrices para la indexación de datos de investigación en repositorios digitales vinculados a estas tres naciones, buscando contribuir a la formalización de procedimientos de indexación más estructurados y transparentes. La investigación adopta un enfoque aplicado, cualitativo, descriptivo y exploratorio, con recolección de datos bibliográficos y documentales. El universo de la investigación estuvo compuesto por repositorios digitales identificados en el directorio re3data. El análisis de documentos provenientes de 21 (veintiún) repositorios calificados para la investigación reveló que ninguno de estos sistemas de autoarchivo presenta un documento formal, específicamente titulado "Política de Indexación", que aborde la representación del contenido temático de los datos de investigación. Aunque dicho documento no fue identificado, podría servir como base para el desarrollo de materiales complementarios, como Manuales de Indexación, Guías de Autoarchivo e Instrucciones para la Asignación de Palabras Clave. Con el objetivo de cumplir con el propósito de la investigación, se identificaron y analizaron 15 (quince) criterios orientadores para la asignación de palabras clave, definidos con base en el marco teórico de la investigación y las observaciones de los documentos analizados. Se constató que las directrices de los repositorios examinados no abordan de manera completa todos los aspectos esenciales para una asignación eficaz de palabras clave por parte de los autores científicos. Las instrucciones proporcionadas suelen ser superficiales y carecen de profundidad y relevancia, con documentos que no presentan procedimientos suficientemente estructurados y claros. La conclusión del estudio señala que la mejora de la calidad de los documentos orientadores para la indexación, especialmente en las prácticas de autoarchivo, debe ser una prioridad para las instituciones responsables de la gestión de estos repositorios. Los criterios presentados en esta investigación pueden servir como base para la creación de documentos más eficientes y sólidos, capaces de orientar de manera más precisa a los autores científicos en la representación temática de los datos de sus investigaciones.

**Palabras clave:** Indexación. Datos de investigación. Palabras clave. Autoarchivo. Repositorios digitales.

## LISTA DE QUADROS

<b>Quadro 1</b> - Tipologias de dados de pesquisa.....	31
<b>Quadro 2</b> - Funções de um sistema de gestão de dados de pesquisa.....	33
<b>Quadro 3</b> - Os 8 Princípios dos Dados Governamentais Abertos.....	37
<b>Quadro 4</b> - Princípios de dados FAIR.....	40
<b>Quadro 5</b> - Tipologias de repositórios de dados de pesquisa.....	46
<b>Quadro 6</b> - Padrões de metadados por área do conhecimento.....	51
<b>Quadro 7</b> - Repositórios de dados de pesquisa do Brasil, Espanha e Portugal que adotam o autoarquivamento (re3data).....	76
<b>Quadro 8</b> - Critérios para a atribuição de palavras-chave.....	83
<b>Quadro 9</b> - Mapeamento dos critérios para a atribuição de palavras-chave na pesquisa documental.....	86

## LISTA DE FIGURAS

<b>Figura 1</b> - Padrões de metadados mais utilizados em repositórios de dados de pesquisa.....	50
<b>Figura 2</b> – Vocabulários controlados e níveis de controle de vocabulário.....	57
<b>Figura 3</b> - Tipologias dos repositórios de dados de pesquisa.....	82
<b>Figura 4</b> - Quantidade de critérios contemplados por repositório de dados de pesquisa.....	88
<b>Figura 5</b> - Nível de prevalência dos critérios analisados na pesquisa documental...	90

## LISTA DE ABREVIATURAS E SIGLAS

ABNT	Associação Brasileira de Normas Técnicas
AHEAD	<i>European Archive of Historical Earthquake Data</i>
AIP	<i>Archival information package</i>
BDTD	Biblioteca Digital Brasileira de Teses e Dissertações
BOAI	<i>Budapest Open Access Initiative</i>
BR	Brasil
BRAPCI	Base de Dados Referencial de Artigos de Periódicos em Ciência da Informação
C&T	Ciência e Tecnologia
CAPES	Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
CF	<i>Climate and Forecast</i>
CNPQ	Conselho Nacional de Desenvolvimento Científico e Tecnológico
CONARQ	Conselho Nacional de Arquivos
CSDGM	<i>Content Standard for Digital Geospatial Metadata</i>
CSV	<i>Comma-separated values</i>
CSUC	<i>Consorti de Serveis Universitaris de Catalunya</i>
DDI	<i>Data Documentation initiative</i>
DFG	<i>German Research Foundation</i>
DIF	<i>Directory Interchange Format</i>
DOI	<i>Digital Object Identifier</i>
DTL	<i>Dutch Techcentre for Life Sciences</i>
EML	<i>Ecological Metadata Language</i>
ENANCIB	Encontro Nacional de Pesquisa em Ciência da Informação
ES	Espanha
FAIR	<i>Findable, Accessible, Interoperable and Reusable</i>
FCT	Fundação para a Ciência e a Tecnologia
FECYT	<i>Fundación Española para la Ciencia y la Tecnología</i>
FINEP	Financiadora de Estudos e Projetos
FGDC/CSDGM	<i>Federal Geographic Data Committee Content Standard for Digital Geospatial Metadata</i>
HEFCE	<i>Higher Education Funding Council For England</i>
IBICT	Instituto Brasileiro de Informação em Ciência e Tecnologia
INCT	Centro de Estudos Integrados da Biodiversidade Amazônica
ISO	<i>International Organization for Standardization</i>
LARHUD	Laboratório em Rede de Humanidades Digitais
LISA	<i>Library &amp; Information Science Abstracts</i>
MCTI	Ministério da Ciência, Tecnologia e Inovação
MEDLINE	Sistema Online de Busca e Análise de Literatura Médica
MeSH	<i>Medical Subject Headings</i>
METS	<i>Metadata Encoding &amp; Transmission Standard</i>

NISO	<i>National Information Standards Organization</i>
NLP	<i>Natural Language Processing</i>
NSF	<i>National Science Foundation</i>
OAI	<i>Open Archives Initiative</i>
OAI-PMH	<i>Open Archives Initiative Protocol for Metadata Harvesting</i>
OAI-ORE	<i>Open Archives Initiative Object Reuse and Exchange</i>
OAIS	<i>Open Archival Information System</i>
OECD	<i>Organization for Economic Co-operation and Development</i>
OCDE	<i>Organização para a Cooperação e Desenvolvimento Econômico</i>
OMS	<i>Organização Mundial de Saúde</i>
ONU	<i>Assembleia Geral das Nações Unidas</i>
ORCID ID	<i>Open Researcher and Contributor ID</i>
PCAL	<i>Portal de Publicação Seriada de Acesso Livre</i>
PDF	<i>Portable Document Format</i>
PID	<i>Persistent Identifier</i>
PREMIS	<i>Preservation Metadata: Implementation Strategies</i>
PT	<i>Portugal</i>
RCAAP	<i>Repositórios Científicos de Acesso Aberto de Portugal</i>
RDF	<i>Resource Description Framework</i>
re3data	<i>Registry of Research Data Repositories</i>
RI	<i>Repositório Institucional</i>
SARDC	<i>Serviço de Alojamento de Repositórios de Dados Científicos</i>
SciELO	<i>Scientific Electronic Library Online</i>
SGML	<i>Standard Generalized Markup Language</i>
SIOR	<i>Social Impact of Science</i>
SOC	<i>Sistemas de Organização do Conhecimento</i>
SPARQL	<i>Protocol and RDF Query Language</i>
SPD	<i>Secretaria de Pesquisa e Desenvolvimento</i>
TRAC	<i>Trustworthy Repositories Audit &amp; Certification</i>
UC	<i>Universidad de Cantabria</i>
UFSCar	<i>Universidade Federal de São Carlos</i>
UNESCO	<i>United Nations Educational, Scientific and Cultural Organization</i>
UNF	<i>Universal Numeric Fingerprint</i>
URL	<i>Uniform Resource Locator</i>
URN	<i>Uniform Resource Name</i>
URV	<i>Universitat Rovira i Virgili</i>
USP	<i>Universidade de São Paulo</i>
W3C	<i>World Wide Web Consortium</i>
XDBD	<i>Expansão da Biblioteca Digital Brasileira</i>
XLS	<i>Excel Spreadsheet File</i>
XML	<i>Extensible Markup Language</i>

## SUMÁRIO

<b>1 INTRODUÇÃO.....</b>	<b>16</b>
1.1 PROBLEMA DE PESQUISA.....	18
1.2 OBJETIVOS.....	18
1.3 JUSTIFICATIVAS.....	19
1.4 ESTRUTURA DA PESQUISA.....	20
<b>2 FUNDAMENTAÇÃO TEÓRICA.....</b>	<b>22</b>
2.1 A CIÊNCIA ABERTA.....	22
2.2 DADOS DE PESQUISA.....	30
2.2.1 Dados de pesquisa abertos.....	35
2.3 REPOSITÓRIOS DIGITAIS.....	43
2.3.1 Repositórios de dados de pesquisa.....	45
2.3.1.1 <i>A prática do autoarquivamento</i> .....	52
2.4 REPRESENTAÇÃO TEMÁTICA DA INFORMAÇÃO E INDEXAÇÃO.....	55
2.5 PALAVRAS-CHAVE NA REPRESENTAÇÃO TEMÁTICA DA INFORMAÇÃO.....	62
2.6 DIRETRIZES E POLÍTICAS PARA ATRIBUIÇÃO DE TERMOS DESCRITORES E PALAVRAS-CHAVE EM SISTEMAS DE AUTOARQUIVAMENTO.....	65
<b>3 PROCEDIMENTOS METODOLÓGICOS.....</b>	<b>70</b>
3.1 CARACTERIZAÇÃO DA PESQUISA.....	70
3.2 UNIVERSO DA PESQUISA.....	71
3.3 PROCEDIMENTO PARA ANÁLISE DOS DADOS.....	83
<b>4 RESULTADOS E DISCUSSÃO.....</b>	<b>85</b>
<b>5 CONSIDERAÇÕES FINAIS.....</b>	<b>97</b>
<b>REFERÊNCIAS.....</b>	<b>102</b>
<b>APÊNDICE A – Triagem de repositórios de dados de pesquisa para a composição do universo de pesquisa (re3data).....</b>	<b>113</b>

## 1 INTRODUÇÃO

A explosão informacional impulsionada pelo avanço da *internet* revelou que, nesse ambiente dinâmico e complexo, objetos informacionais digitais precisam ser organizados de forma eficaz para cumprir seu objetivo primordial: informar. Nesse contexto, a Organização do Conhecimento destaca-se como um subdomínio central da Ciência da Informação, dedicado à construção de sistemas formalizados para a organização conceitual do conhecimento. O ato de *representar* e *organizar* informações em domínios específicos favorece sua recuperação e utilização.

No campo científico, a Ciência Aberta emerge como um movimento global voltado para tornar o conhecimento científico acessível e compartilhado, tanto para a comunidade científica quanto para a sociedade em geral. Além da disponibilização gratuita dos resultados de pesquisas científicas em diferentes plataformas, a abertura de dados brutos relacionados a essas pesquisas tem quebrado barreiras e conquistado ampla aceitação na comunidade científica ao longo das últimas décadas. Esse compartilhamento promove a reutilização dos dados, fomenta a cooperação entre diferentes partes interessadas e acelera o avanço do conhecimento, resultando em benefícios sociais ampliados.

Nesse cenário, a criação de repositórios digitais, fundamentada em uma infraestrutura tecnológica robusta, apresenta-se como uma solução viável e recomendada para o compartilhamento aberto de dados de pesquisa em todas as áreas do conhecimento. Esses repositórios visam aumentar a eficiência na resposta às demandas científicas. A curadoria de dados de pesquisa, que inclui etapas realizadas nesses repositórios, tem como objetivo assegurar a qualidade, a integridade, o acesso e o reuso dos dados armazenados. Entre essas etapas, a indexação destaca-se como o processo técnico essencial para viabilizar a recuperação eficiente de informações e dados. Mais do que simplesmente armazenar objetos digitais, é fundamental garantir condições para seu acesso. No entanto, uma indexação inadequada pode gerar distorções nos resultados das buscas, dificultando o atendimento das necessidades informacionais dos usuários.

Indexação e *recuperação* estão intimamente conectadas. A eficácia de um sistema de recuperação de informações depende da consistência da indexação realizada (Cesarino, 1985). Essa, por sua vez, deve ser respaldada por instrumentos utilizados na etapa de organização da informação (Lima; Álvares, 2012). Diretrizes claras e bem fundamentadas para a representação temática da informação são indispensáveis para resultados satisfatórios, refletindo princípios e orientações que influenciam diretamente a consistência temática e a eficiência dos sistemas de recuperação. Essas diretrizes são de interesse de diversos públicos, incluindo gestores de sistemas, criadores de registros bibliográficos, pesquisadores e autores científicos.

Este estudo, portanto, foca na análise da qualidade e suficiência das diretrizes publicadas em sistemas de informação, tanto em políticas específicas de indexação quanto em outros documentos norteadores do processo, para que autores científicos depositem e indexem seus dados em repositórios digitais, mediante uma prática conhecida como *autoarquivamento*. Indexar dados de pesquisa consiste na atribuição de palavras-chave que reflitam o conteúdo temático dos dados, tornando essas palavras o principal meio de acesso durante a recuperação informacional.

Nos últimos anos, a indexação de dados de pesquisa tem sido objeto de estudo no Grupo de Pesquisa “Organização do Conhecimento e Humanidades Digitais” do Programa de Pós-Graduação em Ciência da Informação da Universidade Federal de São Carlos (UFSCar). Exemplos incluem a pesquisa de Silva (2022), intitulada “*Atribuição de palavras-chave na prática do autoarquivamento: estudo das diretrizes de repositórios de dados de pesquisa ibero-americanos*”, e a de Viana (2020), “*Representação colaborativa de dados científicos: estudo na rede de repositórios de dados científicos do estado de São Paulo*”. Esses estudos, assim como o presente, refletem o interesse pela complexidade pragmática do compartilhamento aberto e da representação de dados de pesquisa em sistemas digitais contemporâneos que adotam o autoarquivamento.

## 1.1 PROBLEMA DE PESQUISA

Sistemas de informação que adotam o autoarquivamento de dados de pesquisa devem incluir em suas diretrizes de indexação um conjunto claro e bem estruturado de critérios orientadores. Pesquisadores da Ciência da Informação, em uma abordagem amplamente reconhecida, sugerem que tais diretrizes e seus critérios associados sejam fundamentados em um documento primário denominado *Política de Indexação*. Esse documento, caracterizado por sua abrangência, articulação e consistência, deve apresentar como outros atributos fundamentais unidade e a completude de informações.

Estudos recentes têm apontado a ausência de políticas que ao menos parte de seu conteúdo guie a representação temática de dados de pesquisa em sistemas de autoarquivamento. Um exemplo é a pesquisa de Silva (2022), conduzida no contexto dos repositórios de dados de pesquisa ibero-americanos, que evidencia essa lacuna. Diante desse problema, crítico tanto para a representação quanto para a recuperação da informação nesses sistemas, esta pesquisa busca responder à seguinte questão investigativa:

*Como as diretrizes de repositórios de dados de pesquisa vinculados ao Brasil, Espanha e Portugal orientam os autores científicos na atribuição de palavras-chave durante o processo de autoarquivamento?*

## 1.2 OBJETIVOS

A presente pesquisa tem como *objetivo geral* analisar critérios norteadores para indexação em diretrizes destinadas ao depósito de dados de pesquisa por autores científicos em repositórios digitais vinculados ao Brasil, Espanha e Portugal. Busca-se, por meio dos resultados obtidos, contribuir para a formalização de procedimentos de indexação mais estruturados e transparentes.

Para alcançar este objetivo, foram definidos os seguintes objetivos específicos:

- a) Evidenciar o papel dos dados de pesquisa e dos repositórios de dados de pesquisa no contexto da Ciência Aberta;
- b) Relacionar o processo de indexação, políticas, diretrizes e critérios para a indexação e a prática do autoarquivamento em sistemas de recuperação de informação;
- c) Identificar repositórios de dados de pesquisa vinculados ao Brasil, Espanha e Portugal que atuam como sistemas de autoarquivamento;
- d) Localizar, nos repositórios digitais selecionados, diretrizes para a indexação de dados de pesquisa pelo autor científico;
- e) Analisar, com base nas recomendações da literatura especializada, os critérios norteadores para a prática da indexação realizada pelo autor científico; e
- f) Apresentar considerações que contribuam para a formalização de procedimentos mais estruturados e transparentes no âmbito da indexação de dados de pesquisa em sistemas de autoarquivamento.

### 1.3 JUSTIFICATIVAS

Identificar e avaliar os esforços dedicados à formulação de diretrizes para a indexação de dados de pesquisa em sistemas de autoarquivamento representa um avanço significativo nas discussões sobre Ciência Aberta. Os resultados obtidos podem estimular novas iniciativas, contribuindo para a adequação e o desenvolvimento de diretrizes e políticas de indexação mais eficazes.

A realização de uma pesquisa que aborde repositórios de dados de pesquisa funcionando como sistemas de autoarquivamento é justificada pela crescente relevância do compartilhamento e do reuso de dados no cenário científico atual, além do aumento expressivo da prática de autoarquivamento nos últimos anos.

A escolha dos países considerados neste estudo reflete o comprometimento dessas nações com a Ciência Aberta por meio de iniciativas relevantes. No Brasil, o Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT) tem desempenhado um papel central como órgão nacional de informação e unidade de pesquisa pública, desenvolvendo desde o início dos anos 2000 infraestruturas, projetos e ações para consolidar os princípios do Acesso Aberto e da Ciência Aberta. Na Espanha, a *Fundación Española para la Ciencia y la Tecnología* (FECYT) tem se destacado pela formulação de políticas otimizadas de Ciência Aberta, implementando-as de maneira eficaz. Em Portugal, a *Fundação para a Ciência e a Tecnologia* (FCT) exerce um papel essencial como entidade promotora do sistema científico e tecnológico nacional, utilizando instrumentos que incentivam e promovem a Ciência Aberta.

Espera-se que os achados deste estudo contribuam para um progresso conjunto entre esses países, fortalecendo o debate e promovendo ações relacionadas à indexação de dados de pesquisa em sistemas de autoarquivamento.

#### 1.4 ESTRUTURA DA PESQUISA

Além desta seção introdutória, esta monografia está estruturada em mais quatro seções principais:

- **Seção 2 – Fundamentação Teórica:** apresenta as bases teóricas que sustentam a pesquisa, situando-a no campo da Biblioteconomia e da Ciência da Informação.
- **Seção 3 – Procedimentos Metodológicos:** descreve os métodos aplicados que permitiram o desenvolvimento da pesquisa de maneira prática e direcionada.
- **Seção 4 – Resultados e Discussão:** inicialmente, mapeia, sistematiza e analisa os critérios orientadores de indexação presentes nas diretrizes de repositórios de dados de pesquisa que adotam o autoarquivamento nos contextos do Brasil, Espanha e Portugal. Em seguida, oferece considerações

que podem colaborar para a formalização de procedimentos mais estruturados e transparentes nesse cenário.

- **Seção 5 – Considerações Finais:** encerra a monografia com as principais conclusões alcançadas ao longo da pesquisa.

## 2 FUNDAMENTAÇÃO TEÓRICA

Esta seção apresenta os resultados da pesquisa bibliográfica sobre os temas abordados.

A subseção 2.1 trata do contexto da *Ciência Aberta* e das práticas que favorecem o acesso aberto à comunicação científica. Na subseção 2.2, são discutidos os *dados de pesquisa* e os *dados de pesquisa abertos* (2.2.1), destacando sua relevância na atualidade. A subseção 2.3 enfoca os *repositórios digitais* no contexto da *Ciência Aberta*, com ênfase nos *repositórios de dados de pesquisa* como tecnologias fundamentais para o armazenamento, preservação e acesso a dados de pesquisas científicas (2.3.1); além disso, aborda a prática do *autoarquivamento* e suas implicações na representação temática da informação (2.3.1.1). Na subseção 2.4, discute-se a *representação temática da informação* e a *indexação*. A subseção 2.5 analisa o papel das *palavras-chave* na representação temática da informação, enquanto, finalmente, a subseção 2.6 aborda questões relacionadas às *diretrizes* para a atribuição de termos descritores e palavras-chave a dados de pesquisa em sistemas de autoarquivamento, as *Políticas de Indexação* e suas relações com diferentes documentos que respaldam esse processo.

### 2.1 A CIÊNCIA ABERTA

Historicamente, o acesso à informação científica passou a ser visto como essencial para a construção de novos conhecimentos e para o progresso da sociedade. Na chamada Sociedade da Informação, surgiu a necessidade de revisar políticas, decisões e práticas que restringiam o conhecimento aberto em diversos níveis e contextos sociais.

De acordo com o Artigo 27 da Declaração Universal dos Direitos Humanos (1948), a *Ciência Aberta* (ou *Open Science*, em inglês) é uma ferramenta crucial para reduzir desigualdades entre países e promover o direito humano de acessar e se beneficiar do progresso científico. Esse Artigo reforça que os benefícios são

potencializados quando as restrições à Ciência são mínimas, especialmente em contextos de financiamento público, permitindo que todos, tanto a comunidade científica quanto a sociedade, possam aproveitar o conhecimento científico sem custos duplicados.

A principal característica da Ciência Aberta, em comparação com práticas científicas anteriores, é a transparência, reutilização, colaboração e inclusão social na produção científica, conforme destacam Silva e Silveira (2019). Albagli, Clinio e Raychtock (2014) veem a Ciência Aberta como um conceito abrangente, que envolve práticas, abordagens e um novo modelo para projetar, executar e comunicar pesquisas, frequentemente utilizando ferramentas para gestão, armazenamento e acesso à informação.

Neste sentido, o movimento de Ciência Aberta está associado a um conjunto de inovações relativamente recentes, associadas às novas formas de colaboração e culturas de uso em ambientes digitais; mais especificamente, à disseminação da chamada *cultura livre digital*, que oferece suporte desde a concepção da investigação até a comunicação e a divulgação científica com o uso de *softwares* abertos. Para Belli *et al.* (2023, p. 12), se podemos falar de Ciência Aberta ou de Acesso Aberto, “[...] isso deve-se precisamente às possibilidades oferecidas pela *internet*, da mesma forma que a maior produção científica está associada à existência de plataformas tecnológicas de divulgação”.

A UNESCO, como uma agência das Nações Unidas com mandato para a Ciência, define a Ciência Aberta como um processo inclusivo, que visa aumentar a colaboração científica e o compartilhamento de informações para beneficiar a Ciência e a sociedade. De acordo com a UNESCO (2021), a Ciência Aberta é baseada em quatro pilares principais: Conhecimento científico aberto (*Open scientific knowledge*), Infraestruturas de Ciência Aberta (*Open science infrastructures*), Engajamento aberto de atores sociais (*Open engagement of societal actors*), e Diálogo aberto com outros sistemas de conhecimento (*Open dialogue with other knowledge systems*).

O primeiro pilar, *conhecimento científico aberto*, refere-se ao acesso aberto a publicações científicas, dados de pesquisa, metadados, recursos educacionais

abertos, *software*, código-fonte e *hardware*, que devem estar disponíveis no domínio público ou sob licenças abertas, permitindo acesso, reutilização, adaptação e distribuição de maneira gratuita, sem restrições de localização, nacionalidade ou qualquer outra condição, conforme detalhado pela UNESCO (2021).

O segundo pilar, *infraestruturas de Ciência Aberta*, inclui recursos compartilhados, como repositórios, plataformas de publicação de acesso aberto e sistemas de avaliação e análise de dados científicos, essenciais para o apoio à Ciência Aberta.

O terceiro pilar, *engajamento aberto de atores sociais*, promove a colaboração entre cientistas e outros envolvidos sociais, além da comunidade científica, tornando o processo científico mais inclusivo e acessível. A UNESCO (2021) também destaca novas formas de colaboração, como *crowdfunding* e voluntariado científico.

O quarto pilar, *diálogo aberto com outros sistemas de conhecimento*, implica no diálogo entre diversos detentores de conhecimento, reconhecendo a riqueza de diferentes sistemas e epistemologias, alinhado à Declaração Universal sobre Diversidade Cultural da UNESCO (2001).

A UNESCO também estabelece os *Valores da Ciência Aberta*, que envolvem:

- *Qualidade e integridade (Quality and integrity)*, que reconhece o respeito à liberdade acadêmica e aos direitos humanos na execução da Ciência Aberta;
- *Benefício coletivo (Collective benefit)*, que reconhece que a Ciência Aberta como pertencente à humanidade e deve beneficiar a comunidade como um todo;
- *Equidade e justiça (Equity and fairness)*, que reconhece que a Ciência Aberta deve desempenhar um papel significativo em garantir a equidade entre pesquisadores de países desenvolvidos e em desenvolvimento; e
- *Diversidade e inclusão (Diversity and inclusiveness)*, que reconhece que a Ciência Aberta deve abranger uma diversidade de conhecimentos, práticas, fluxos de trabalho, linguagens, resultados de pesquisa e tópicos de pesquisa que apoiem as necessidades e o pluralismo epistêmico da comunidade científica como um todo, incluindo detentores de conhecimento para além da

comunidade científica tradicional, como povos indígenas e comunidades locais, e atores sociais de diferentes países e regiões, conforme apropriado.

Esses valores garantem que a Ciência Aberta beneficie a comunidade global de maneira ética e justa, apoiando a inclusão de comunidades marginalizadas.

Além disso, os *Princípios da Ciência Aberta*, segundo a UNESCO (2021), asseguram os Valores da Ciência Aberta e incluem:

- *Transparência, escrutínio, crítica e reprodutibilidade (Transparency, scrutiny, critique and reproducibility)*, com o entendimento de que, uma maior abertura deve ser promovida em todas as fases do esforço científico para aumentar o impacto social da ciência e aumentar a capacidade da sociedade como um todo para resolver problemas complexos interligados. De que, uma maior abertura leva a uma maior transparência e confiança na informação científica e reforça a característica fundamental da ciência como uma forma distinta de conhecimento baseada em evidências e testada contra a realidade, a lógica e o escrutínio de pares científicos;
- *Igualdade de oportunidades (Equality of opportunities)*, com o entendimento de que, todos os cientistas, outros atores e partes interessadas da Ciência Aberta, independentemente de localização, nacionalidade, raça, idade, gênero, renda, circunstâncias socioeconômicas, estágio de carreira, disciplina, idioma, religião, deficiência, etnia ou *status* migratório, têm a mesma oportunidade de acessar, contribuir e se beneficiar da Ciência Aberta;
- *Responsabilidade, respeito e prestação de contas (Responsibility, respect and accountability)*, com o entendimento de que, uma maior abertura acarreta maior responsabilidade para todos os atores da Ciência Aberta. De que, a responsabilidade pública, a sensibilidade aos conflitos de interesse, a vigilância quanto às possíveis consequências sociais e ecológicas das atividades de pesquisa, a integridade intelectual e o respeito pela ética, princípios e implicações relativos à pesquisa, devem formar a base para a boa governança da Ciência Aberta;

- *Colaboração, participação e inclusão (Collaboration, participation and inclusion)*, com o entendimento de que, colaborações em todos os níveis do processo científico, além das fronteiras da geografia, língua, gerações, disciplinas e recursos, devem se tornar a norma, juntamente com a participação plena e efetiva dos atores sociais e inclusão de conhecimento de comunidades marginalizadas na resolução de problemas de importância social;
- *Flexibilidade (Flexibility)*, com o entendimento de que, devido à diversidade de sistemas científicos, atores e capacidades em todo o mundo, bem como a natureza evolutiva das tecnologias de informação e comunicação de apoio, não existe uma forma única de praticar a Ciência Aberta. Diferentes caminhos de transição e prática da Ciência Aberta precisam ser encorajados, mantendo os valores centrais acima mencionados e maximizando a adesão aos outros princípios aqui apresentados; e, por fim,
- *Sustentabilidade (Sustainability)*, com o entendimento de que, para ser o mais eficiente e impactante possível, a Ciência Aberta deve se basear em práticas, serviços, infraestruturas e modelos de financiamento de longo prazo que garantam a participação igual de produtores científicos de instituições e países menos privilegiados. As infraestruturas de Ciência Aberta devem ser organizadas e financiadas por uma visão de longo prazo sem fins lucrativos, que aprimore as práticas de Ciência Aberta e garanta o acesso permanente e irrestrito a todos na medida do possível.

Esses princípios visam garantir a boa governança, inclusão, participação e continuidade das práticas de Ciência Aberta em nível global.

Em 2002, a UNESCO já havia publicado a Declaração de Budapeste, que promoveu uma mudança nos habituais modelos de armazenamento, disseminação e acesso à produção científica, se voltando ao *oligopólio das editoras científicas* - circunstância de mercado em que poucas empresas detêm o controle da maior parcela do mesmo. Essa declaração introduziu a Via Verde (*Green Road*), que incentiva o autoarquivamento, sendo essa uma prática em que os próprios autores

submetem seus trabalhos para um espaço digital denominado repositório digital; e a Via Dourada (*Golden Road*), que visou a criação de revistas de acesso aberto na *web*.

Com perspectiva similar, em 2003, foram criadas a “Declaração de Bethesda sobre Publicação em Acesso Aberto”, oriunda da Reunião de Bethesda, que se direcionou às preocupações da área biomédica para a Ciência Aberta; e a “Declaração de Berlim sobre o Acesso Aberto ao Conhecimento nas Ciências e Humanidades”.

A Reunião de Bethesda estabeleceu uma definição para “publicações de acesso aberto”, contemplando dois pontos cruciais. Primeiro, considerando a problemática dos direitos autorais, determinou que, no contexto do acesso aberto, os autores devem assegurar aos usuários o direito livre, irrevogável e perpétuo de acesso a seus trabalhos, incluindo

[...] 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 estabeleceu a obrigatoriedade do depósito integral dos trabalhos e materiais relacionados à pesquisa, juntamente com uma cópia da licença para uso e distribuição, em um repositório digital. Esse repositório deve estar vinculado a uma instituição acadêmica, científica ou governamental, garantindo acesso livre imediato e interoperabilidade entre sistemas, além de arquivamento de longo prazo.

Reconhecendo a crescente importância da *internet* e as discussões prévias sobre a necessidade do acesso aberto, a *Declaração de Berlim* (2003) definiu duas condições essenciais para o acesso aberto:

- 1) O autor e o detentor dos direitos de tais contribuições concedem para todos os utilizadores/usuários o direito livre e gratuito, irrevogável e mundial de acessar a obra e licenciam a sua cópia, uso, distribuição, transmissão e disposição pública e a elaboração e distribuição de obras derivadas em qualquer meio digital para qualquer propósito responsável, sujeito à atribuição adequada de

autoria, assim como o direito de fazer poucas cópias para o seu uso pessoal; e

2) A versão completa do trabalho e todos os materiais complementares, incluindo a cópia da permissão supracitada (e portanto publicada) é depositada em formato eletrônico padrão em ao menos um repositório usando padrões técnicos adequados (tais como as definições do *Open Archive*) que é mantido por uma instituição acadêmica, sociedade científica, agência governamental ou outra instituição bem estabelecida que busca permitir o acesso aberto, a distribuição irrestrita, a interoperabilidade e o arquivamento de longo prazo. (*Berlin Declaration On Open Access To Knowledge In The Sciences And Humanities*, 2003).

Em 2005, uma conferência subsequente refinou a Declaração de Berlim, resultando na definição de dois princípios-chave:

- os signatários devem exigir que os pesquisadores depositem uma cópia de seus trabalhos em um repositório de acesso aberto; e
- os signatários devem encorajar a publicação do trabalho em periódicos de acesso aberto, quando disponíveis.

Esses princípios, referenciados como "Green OA" e "Gold OA", respectivamente, formam o chamado *mandato de acesso aberto*. Em 2013, no 10º aniversário da Declaração de Berlim, mediante uma Declaração de Missão, foi estabelecida a meta de garantir que 90% das pesquisas fossem publicadas em acesso aberto.

Contudo, a "Iniciativa dos Arquivos Abertos" (*Open Archives Initiative - OAI*), lançada na Convenção de Santa Fé (1999), é considerada pioneira no movimento de acesso aberto. Essa iniciativa formalizou o protocolo OAI-PMH (*Open Archives Initiative Protocol for Metadata Harvesting*), promovendo normas de interoperabilidade para melhorar a recuperação e o compartilhamento de conteúdo científico. De acordo com a *National Information Standards Organization* (NISO, 2004, p. 2), interoperabilidade é a "capacidade de múltiplos sistemas, com diferentes plataformas de *hardware* e *software*, estruturas de dados e interfaces, trocarem dados com mínima perda de conteúdo e funcionalidade".

Segundo Weitzel (2006), a OAI é caracterizada por:

- permitir o autodepósito<sup>1</sup> (*upload*) de produções científicas pelos autores, em múltiplos formatos e com metadados padronizados; e
- garantir acesso livre a essas produções, com possibilidade de *download*, aplicando o protocolo OAI-PMH.

No Brasil, o IBICT tem liderado as articulações para a implantação do movimento de livre acesso à informação desde os anos 2000, criando infraestruturas, projetos e ações para concretizar os preceitos defendidos pelo Acesso Aberto e pela Ciência Aberta. Em 2005, publicou o “Manifesto Brasileiro de Apoio ao Acesso Livre à Informação Científica”, que destacou as dificuldades enfrentadas pela comunidade científica no acesso à informação. Em 2009, o IBICT lançou duas iniciativas importantes:

1. um projeto-piloto para o desenvolvimento de repositórios institucionais nacionais; e
2. o “Edital IBICT-FINEP/PCAL/XDBD nº 002/2009”, que distribuiu *kits* tecnológicos e suporte para implementar o *software* de código aberto DSpace<sup>2</sup>, proporcionando maior visibilidade científica às instituições contempladas.

Na Espanha, a *Fundación Española para la Ciencia y la Tecnología* (FECYT), uma organização pública, destaca-se na formulação e implementação otimizada de políticas voltadas para a Ciência Aberta, promovendo avanços significativos nessa área. Em Portugal, esforços semelhantes têm sido conduzidos pela *Fundação para a Ciência e a Tecnologia* (FCT), uma entidade de grande relevância no fomento do sistema científico e tecnológico nacional. A FCT adota instrumentos que incentivam práticas de Ciência Aberta no contexto das atividades científicas que financia, consolidando-se como um agente estratégico na promoção dessa agenda.

---

<sup>1</sup> Referenciado com frequência como “autoarquivamento”.

<sup>2</sup> Disponível em: <https://dspace.lyrasis.org/download/>.

Inserido no escopo da Ciência Aberta, o conceito de Acesso Aberto refere-se à disponibilidade gratuita e irrestrita dos resultados de pesquisas científicas para qualquer pessoa, facilitada por meio de plataformas tecnológicas como repositórios digitais. Além disso, abrange diversos objetos informacionais associados às pesquisas, incluindo as fontes primárias, ou seja, os dados brutos que sustentam as investigações científicas. Esses aspectos serão explorados mais detalhadamente na próxima seção.

## 2.2 DADOS DE PESQUISA

Dados de pesquisa são a base para se debater o conhecimento. Segundo o *Higher Education Funding Council For England* (HEFCE) e associados (2016), eles correspondem às informações que respondem à pergunta central de um pesquisador. Curty *et al.* (2017, p. 1) afirmam que “dados de pesquisa são a espinha dorsal da descoberta científica e da inovação tecnológica, sendo considerados a moeda principal da ciência e os pilares da pesquisa”.

No contexto brasileiro, uma definição amplamente aceita é apresentada por Sales e Sayão (2019, p. 36), que caracterizam dados de pesquisa como “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”.

Esses dados, fontes primárias da investigação científica, assumem papel estratégico e valioso no avanço da Ciência. Antes vistos como subprodutos da pesquisa, no contexto da Ciência Aberta, passaram a ser amplamente reconhecidos como recursos reutilizáveis, com aplicações além das previstas pelos pesquisadores que os geraram, despertando grande interesse em universidades, agências de fomento e outras instituições que atuam como produtoras e consumidoras desses dados.

Presentes em todas as áreas do conhecimento, os dados de pesquisa variam conforme os contextos em que são produzidos. Eles podem ser coletados ou gerados de diversas formas, e sua classificação depende de características como tipo, complexidade e finalidade, o que influencia decisões sobre sua preservação.

Kitchin (2014, p. 21) ressalta a importância dos dados no mundo contemporâneo, definindo-os como “a matéria-prima produzida pela abstração do mundo em categorias, medidas e outras formas representacionais – números, caracteres, símbolos, imagens, sons, ondas eletromagnéticas, *bits* [...]”. Para o autor, os dados constituem a base da informação e do conhecimento, e, mesmo isolados, possuem valor representativo. Muitos dados podem gerar outros dados, sendo registrados em formatos analógicos ou digitais.

A *National Science Foundation*<sup>3</sup> (NSF, 2020) propõe tipologias para os dados de pesquisa considerando critérios como o *procedimento de coleta*, *caráter*, *grau de estruturação*, *nível de abertura* e *formato*. Essas tipologias são apresentadas no Quadro 1.

**Quadro 1** - Tipologias de dados de pesquisa.

Critérios	Tipos	Descrição
Quanto ao procedimento de coleta	Observacionais	São procedentes de pesquisas científicas; possuem caráter único, pois não se pode voltar a reproduzi-los; são capturados em tempo real e geralmente fora de laboratório; são registros de fatos ou evidências de fenômenos.
	Computacionais	São procedentes da execução de modelos computacionais, simulações ou fluxos de trabalho; são reproduzíveis se preservada a documentação de <i>hardware</i> e <i>software</i> , os dados de entrada e os passos intermediários.
	Experimentais	São procedentes de experimentos, procedimentos realizados em condições controladas, com o fim de provar ou estabelecer hipóteses; caso seja um experimento replicável, os dados são mais fáceis de reutilizar e preservar.
Quanto ao	Primário	Correspondem aos trabalhos originais de pesquisa

<sup>3</sup> Agência federal independente criada pelo Congresso dos Estados Unidos em 1950 com intuito de promoção e progresso da Ciência; da saúde, da prosperidade e bem-estar nacional e para garantia da defesa nacional.

caráter		e/ou a dados brutos sem interpretação coletados a partir de experimentos, pesquisas, entrevistas e demais técnicas; são coletados a partir do problema de pesquisa; têm custo de obtenção maior; geralmente são utilizados nas Ciências, principalmente nas experimentais.
	Secundário	São considerados menos precisos, pois não foram coletados para responder ao problema de pesquisa; correspondem a dados coletados, tabulados, ordenados e disponíveis publicamente em livros, periódicos, censos, biografias, artigos, bases de dados etc.; são mais utilizados nas Ciências Sociais; permitem a repetição dos estudos e a criação de grandes conjuntos de dados, mais ricos e sofisticados.
	Terciário	Correspondem à forma de dados derivados, como recontagem, categorias e resultados de dados estatísticos; frequentemente são utilizados para garantir a confidencialidade dos dados primários e secundários.
Quanto ao grau de estruturação	Estruturados	São facilmente transferidos a outros sistemas devido à organização segundo um modelo definido; apresentam informação armazenada em tabelas e bases de dados relacionais seguindo uma estrutura determinada onde se definem as tabelas, os campos das tabelas e as relações entre ambos.
	Semiestruturados	Tradicionalmente, inclui imagens, documentos de textos e outros objetos que fazem parte de uma base de dados; não têm um modelo de dados ou uma estrutura pré-definida, não sendo possível manter em uma estrutura de base relacional; são irregulares e flexíveis, frequentemente adicionados hierarquicamente; possuem um conjunto consistente de conteúdo semântico e algum meio de classificá-los e ordená-los.
	Não estruturados	Não possuem um modelo definido ou estrutura identificável; cada elemento individual pode ter uma estrutura ou formato; nem todos os dados dentro de um conjunto possuem a mesma estrutura.
Quanto ao nível de abertura	Dados abertos	Podem ser utilizados e distribuídos por qualquer pessoa, sem barreiras técnicas ou legais sob os requisitos de reconhecer a autoria e compartilhar o novo produto nas mesmas condições.
Quanto ao formato	Arquivos digitais	O formato e o <i>software</i> com que se criam os dados dependem da forma que os pesquisadores coletam e analisam e geralmente são influenciados pelas normas e costumes de cada disciplina; apresentam grande quantidade de formatos disponíveis; correspondem a uma opção mais segura para garantir o acesso e o uso de formatos e padrões.

Fonte: *National Science Foundation* (2020).

Os dados de pesquisa percorrem um ciclo de vida que se inicia com sua coleta, realizada por meio de procedimentos e instrumentos específicos; passa pelo depósito, tratamento e gerenciamento, apoiado por estruturas tecnológicas como repositórios digitais; e culmina em sua recuperação, acesso e reutilização. Sayão e Salles (2016) organizam o papel de um sistema de gestão de dados de pesquisa com base nesse ciclo de vida, conforme ilustrado no Quadro 2.

**Quadro 2** - Funções de um sistema de gestão de dados de pesquisa.

<b>CAPTURA DE DADOS</b> Inserção de coleções de dados – primários ou derivados - provenientes de experimentos, simulações, observações, questionários, levantamentos, etc. Os dados podem ser incorporados ao repositório pelos próprios autores - por autossucessão - ou por equipes especializadas vinculadas ao serviço	
<b>FUNÇÕES</b>	Seleção dos dados passíveis de serem arquivados; verificação do enquadramento no escopo do repositório; verificação dos formatos de arquivos aceitáveis para submissão; verificação dos direitos associados às coleções ( <i>copyright</i> e licenças); verificação de dados sensíveis (dados não anonimizados, confidenciais, pessoais); verificação do volume e quantidade de arquivos; verificação dos metadados gerais e disciplinares que acompanham os dados; normalização para elenco de formatos padronizados aceitos para arquivamento e disseminação; controle de qualidade dos dados; definição de tempo de embargo
<b>CATALOGAÇÃO DAS COLEÇÕES DE DADOS</b> Descrição, atribuição de metadados e inclusão de documentação que assegurem que os dados possam ser acessados e interpretados no tempo e no espaço	
<b>FUNÇÕES</b>	Atribuição de: metadados descritivos, estruturais, administrativos, técnicos (que inclui os relativos às dependências técnicas dos objetos digitais); atribuição de metadados de preservação, que assegurem a proveniência, autenticidade e integridade dos dados ao longo do tempo; uso de taxonomias especializadas e disciplinares; atribuição de identificador persistente (DOI, Handles, UNF, URN, etc.) que permita que os dados possam ser localizados de forma persistente e citados como as publicações acadêmicas; identificação do autor (ORCID ID, Scopus Author ID, ResearcherID, etc.); inclusão de documentação sobre os dados, incluindo descrição do projeto, dos arquivos e dos parâmetros; cadernos de laboratório e de campo, protocolos de pesquisa ou metodologia, etc.; Vinculação (por <i>links</i> ) a publicações e a dados relacionados internos e externos ao repositório
<b>ARQUIVAMENTO E PRESERVAÇÃO</b> Arquivamento seguro que garanta a gestão de curto e longo prazo das coleções de dados orientados por um plano/política de preservação digital	
<b>FUNÇÕES</b>	Armazenamento em sistemas seguros; gestão da preservação de curto prazo

	( <i>backups</i> , <i>backups</i> redundantes <i>offsite</i> ; checagem de integridade, armazenamento seguro, criptografia, compressão); gestão de longo prazo (migração, emulação, reformatação para formatos padronizados, aplicação de normas pertinentes (OAIS, TRAC), informação de fixidade voltada para validar a autenticidade e integridade de um objeto digital ( <i>checksums</i> , assinatura digital) Implementação de trilhas de auditoria
INTEROPERABILIDADE Intercâmbio e compartilhamento e <i>linkage</i> com outros repositórios de dados e outros sistemas de informação (repositórios institucionais, bibliotecas digitais de publicações acadêmicas, editoras científicas)	
FUNÇÕES	Disponibilização de metadados segundo o protocolo OAI-PMH; agregação para formação de publicações ampliadas segundo o padrão OAI-ORE; uso dos padrões, <i>web service</i> , <i>linked data</i> e outros; empacotamento de metadados para intercâmbio segundo o padrão METS
RECUPERAÇÃO, ACESSO E REUSO Interface <i>web</i> para a descoberta, acesso e <i>download</i> de coleções de dados relevantes para o usuário ou para aplicações computacionais, como visualização e mapeamento, que podem prover serviços a partir dessas coleções; vinculado a uma política de acesso estabelecida pela instituição que inclui: tempo de embargo, direito de acesso, pagamentos, restrições sobre determinadas coleções, acesso somente aos metadados; registros de usuários e termos de uso dos dados	
FUNÇÕES	Disponibilização de interfaces <i>web</i> para recuperação, acesso e <i>download</i> ; oferta de aplicações e serviços sobre as coleções

Fonte: Sayão e Sales (2016, p. 100–101).

Sant’Ana (2016) observa que o campo científico da Ciência da Informação tem o potencial de propor “[...] novos caminhos para que os usuários, de forma mais democrática e aberta, possam acessar e utilizar dados, reduzindo a assimetria informacional que pode surgir entre os que os detêm e os que deles precisam”. Nesse contexto, o autor oferece uma perspectiva alternativa sobre o ciclo de vida dos dados, definindo quatro fases principais: *Coleta*, *Armazenamento*, *Recuperação* e *Descarte*. O autor destaca que a maioria das investigações na Ciência da Informação se concentra nas fases de *Coleta* e *Recuperação*, enquanto as etapas de *Armazenamento* e *Descarte*, devido à sua natureza técnica, são exploradas com maior profundidade pela Ciência da Computação. Contudo, Sant’Ana (2016) reforça a importância das contribuições da Ciência da Informação para essas fases, mesmo quando apresentam desafios técnicos mais complexos. Além disso, ele identifica fatores fundamentais que permeiam todas as etapas do ciclo de vida dos dados: *Privacidade*, *Integração*, *Qualidade*, *Direitos Autorais*, *Disseminação* e *Preservação*.

O guia *Open Science Research Data*, publicado em 2024 pelo Ministério Francês de Ensino Superior e Pesquisa, com a coordenação editorial da Universidade de Lille e apoio do conselho científico *The Skills and Training College of The Committee for Open Science*, alerta para os impactos da produção descontrolada de dados de pesquisa no ambiente digital. O documento enfatiza a importância da reutilização de dados existentes antes de gerar novos, alinhando práticas de compartilhamento à Ciência Aberta. Também destaca a necessidade de destruir dados inadequadamente descritos ou redundantes, considerando que tais práticas ajudam a reduzir a pegada digital dos dados (*Open Data: Research Data*, 2024).

A falta de metadados adequados pode comprometer o valor dos dados, prejudicando sua identificação, confiabilidade e viabilidade para uso e reuso. Assim, práticas adequadas de descrição de dados são essenciais para garantir sua efetividade em contextos de compartilhamento e reutilização, contribuindo para os objetivos da Ciência Aberta.

As questões relacionadas à abertura, compartilhamento e reutilização de dados de pesquisa, bem como a descrição de objetos digitais para recuperação em sistemas de informação, são abordadas em maior detalhe nas próximas seções.

### 2.2.1 Dados de pesquisa abertos

O conceito de *dados abertos* (*Open Data*) refere-se à ideia de que certos dados devem estar disponíveis para utilização, reutilização e redistribuição por qualquer pessoa, sem restrições relacionadas a direitos autorais, patentes ou outros mecanismos de controle. Tais dados devem, no máximo, ser submetidos à exigência de atribuição à fonte original e ao compartilhamento sob as mesmas licenças em que foram acessados.

O movimento de Dados Abertos é frequentemente associado a outras iniciativas "abertas", como os movimentos de Código Aberto<sup>4</sup>, Conteúdo Aberto<sup>5</sup> e Acesso Aberto, compartilhando objetivos comuns de promover o livre acesso à informação e fomentar a colaboração.

A filosofia de abertura de dados tem raízes profundas. Por exemplo, em 1955, o Conselho Internacional das Uniões Científicas<sup>6</sup> criou centros de dados globais para minimizar o risco de perda de dados e maximizar sua acessibilidade, recomendando que esses dados fossem disponibilizados em formatos legíveis por máquina. No entanto, o termo "dados abertos" em si é relativamente recente, ganhando força com o crescimento da *internet* e da *World Wide Web*.

Atualmente, as principais fontes de dados abertos são os setores *científico*, *governamental*, *empresarial privado* e *acadêmico*. Cada um desses setores pode ter suas próprias diretrizes e princípios para disponibilizar dados abertos, levando em consideração suas especificidades.

A compreensão de que dados abertos governamentais são essenciais para o fortalecimento da democracia levou à publicação, em 2007, dos "8 Princípios dos Dados Governamentais Abertos" (*8 Principles of Open Government Data*), elaborados por defensores do movimento de Governo Aberto<sup>7</sup> durante um *workshop* realizado em Sebastopol, Califórnia. De acordo com esses princípios, os *dados governamentais abertos* devem ser disponibilizados de maneira a respeitar as seguintes características (Quadro 3):

---

<sup>4</sup> Código Aberto é um movimento que prega a disponibilização de códigos-fonte gratuitamente para consulta, examinação, modificação e redistribuição. Tem suas raízes no movimento do *software livre*, que começou no início dos anos 1980.

<sup>5</sup> Conteúdo Aberto ou Conteúdo Livre corresponde a qualquer tipo de trabalho criativo (por exemplo, artigos, imagens, áudio, vídeo etc.) que pode ser utilizado sem ou com apenas algumas restrições legais.

<sup>6</sup> Na atualidade, denominado Conselho Internacional para a Ciência.

<sup>7</sup> Cultura de governança que promove a colaboração entre governo e sociedade por meio da transparência na gestão, participação social e da responsabilidade e responsividade dos agentes públicos. Tem como objetivo aprimorar políticas e serviços e solucionar questões de interesse público (Gov.br, s/d).

**Quadro 3 - Os 8 Princípios dos Dados Governamentais Abertos.**

Princípio	Fundamento
<i>Complete</i> (Completo)	Dados considerados públicos devem ser disponibilizados. <i>Dados públicos</i> são dados que não estão sujeitos a limitações válidas de privacidade, segurança ou privilégio.
<i>Primary</i> (Primário)	Dados devem ser coletados na fonte, com o maior nível possível de granularidade, não em formas agregadas ou modificadas.
<i>Timely</i> (Oportuno)	Dados devem ser disponibilizados o mais rápido possível para preservar seu valor.
<i>Accessible</i> (Acessível)	Dados devem estar disponíveis para a mais ampla gama de usuários e de finalidades possíveis.
<i>Machine processable</i> (Processável por máquina)	Dados devem ser razoavelmente estruturados para permitir o processamento automatizado.
<i>Non-discriminatory</i> (Não discriminatório)	Dados devem estar disponíveis para qualquer pessoa, sem necessidade de cadastro.
<i>Non-proprietary</i> (Não proprietário)	Dados devem estar disponíveis em um formato sobre o qual nenhuma entidade tem controle exclusivo.
<i>License-free</i> (Sem licença)	Dados não devem estar sujeitos a qualquer regulamento de direitos autorais, patentes, marcas comerciais ou segredos comerciais. Podem ser permitidas restrições razoáveis de privacidade, segurança e privilégios.

Fonte: opengovdata.org (2007).

De acordo com a Public.Resource.Org (s/d), uma organização dedicada à publicação e ao compartilhamento de materiais de domínio público nos Estados Unidos e internacionalmente, a adoção dos oito princípios para disponibilização de dados pode tornar os governos mais eficazes, transparentes e relevantes para a vida de seus cidadãos. A *Organization for Economic Co-operation and Development*<sup>8</sup> (OECD, s/d) reforça essa ideia, afirmando que, ao tornarem seus conjuntos de dados acessíveis, as instituições públicas promovem maior transparência e responsabilidade perante os cidadãos. Além disso, ao incentivar o uso, reutilização e distribuição gratuita desses dados, os governos estimulam a criação de negócios e serviços inovadores, focados no cidadão.

No campo científico, uma das principais iniciativas tem sido garantir o acesso universal aos dados de pesquisa, especialmente aqueles provenientes de estudos

<sup>8</sup> Em português, "Organização para a Cooperação e Desenvolvimento Econômico" (OCDE).

financiados com recursos públicos. Em 2004, a OECD publicou a *Declaration on Access to Research Data from Public Funding*<sup>9</sup>, defendendo que todos os dados de pesquisa financiados com recursos públicos devem ser disponibilizados ao público. Em 2007, a organização publicou os “*OECD Principles and Guidelines for Access to Research Data from Public Funding*”<sup>10</sup>, oferecendo recomendações amplas para políticas científicas governamentais e para órgãos de financiamento dos países membros sobre o acesso a dados de pesquisa.

A abertura de dados de pesquisa é crucial para a reprodutibilidade científica e acelera o avanço do conhecimento. Como afirmam Sayão e Sales (2015, p. 21), “o acesso e compartilhamento de dados de pesquisa contribuem de maneira significativa para que a ciência avance e maximize os investimentos aplicados em programas de pesquisa”. Na mesma linha, Curty *et al.* (2017, p. 1) destacam que,

No passado, os dados de pesquisa custosamente conquistados, eram compartilhados apenas entre alguns colegas confiáveis e conhecidos. Cada vez mais, entretanto, os pesquisadores devem ser responsáveis pelo dinheiro investido por contribuintes e disponibilizar os dados de pesquisa financiadas publicamente, para a comunidade mais ampla, depositando-os em repositórios digitais, por exemplo.

Rauen (2018) lista diversos benefícios associados à abertura de dados de pesquisa, especialmente quando compartilhados em ambientes digitais, como:

- ampliação da colaboração entre pesquisadores da mesma área;
- aumento da visibilidade da pesquisa vinculada aos dados;
- maior eficiência no uso do investimento público em atividades científicas;
- possibilidade de verificação dos resultados das investigações pela comunidade científica;
- democratização e maior transparência nos resultados das pesquisas; e
- aceleração do processo de inovação.

---

<sup>9</sup> Em português, “Declaração sobre acesso a dados de investigação provenientes de financiamento público”.

<sup>10</sup> Em português, “Princípios e Diretrizes da OCDE para Acesso a Dados de Pesquisa de Financiamento Público”.

Em geral, dados de pesquisa disponibilizados de forma aberta em plataformas digitais contêm informações integrais que ainda não passaram por processamento. Para que esses dados sejam reutilizados, é necessário que sejam processados, de modo a gerar novos dados derivados, que serão analisados de acordo com os objetivos do novo pesquisador e sua pesquisa. Contudo, é importante destacar que nem todos os dados de pesquisa são passíveis de abertura. A disponibilização desses dados deve ser acompanhada por uma política de proteção que garanta, por exemplo, o cumprimento de normas de privacidade e segurança (HEFCE *et al.*, 2016).

A partir dos anos 2000, as políticas em favor do acesso aberto aos dados de pesquisa financiados com recursos públicos passaram a ganhar maior ênfase, impulsionando a difusão e o compartilhamento desses dados. Com o reconhecimento dos benefícios do compartilhamento de dados de pesquisa, governos, universidades e instituições de fomento passaram a investir em políticas para promover o acesso aberto, exigindo, por exemplo, a elaboração de planos de gestão de dados de pesquisa e a adesão a diretrizes específicas para o compartilhamento em repositórios digitais (Rauen, 2018).

Nesse contexto, os *Princípios FAIR (Findability, Accessibility, Interoperability, and Reusability*<sup>11</sup>) se destacam como a abordagem mais popular para garantir a qualidade dos dados de pesquisa disponibilizados em sistemas de recuperação da informação. Esses princípios começaram a ser discutidos em 2014, durante a conferência "*Jointly Designing a Data FAIRPORT*", organizada pela instituição holandesa *Elixir*, o *Dutch Techcentre for Life Sciences (DTL)*, o *Netherlands eScience Center* e o *Lorentz Center*. O evento reuniu editores, especialistas, cientistas da computação e instituições de pesquisa para discutir os aspectos necessários à criação de uma infraestrutura global para a publicação, descoberta, compartilhamento e reutilização de dados científicos no contexto das *eScience*<sup>12</sup>. Como resultado, foram definidos os quatro princípios fundamentais para o

---

<sup>11</sup> Em português, "Localização, Acessibilidade, Interoperabilidade, e Reusabilidade".

<sup>12</sup> Refere-se aos métodos de obtenção de resultados científicos por meio da utilização intensiva de computação, usualmente paralela, e de imenso volume de dados.

compartilhamento de dados, que ficaram conhecidos como FAIR. Em 2016, esses princípios foram mais detalhados e apresentados da seguinte forma:

*Para ser localizável:*

- F1. (meta)dados são atribuídos a um identificador globalmente único e persistente;
- F2. os dados são descritos com metadados ricos;
- F3. (meta)dados incluem clara e explicitamente o identificador dos dados que eles descrevem;
- F4. (meta)dados são registrados ou indexados em um recurso pesquisável;

*Para ser acessível:*

- A1. (meta)dados 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:*

- I1. (meta)dados usam uma linguagem formal, acessível, compartilhada e amplamente aplicável para a representação do conhecimento;
- I2. (meta)dados usam vocabulários que seguem os princípios FAIR;
- I3. (meta)dados incluem referências qualificadas a outros (meta)dados;

*Para ser reutilizável:*

- R1. (meta)dados são ricamente descritos com uma pluralidade de atributos precisos e relevantes;
- R1.1. (meta)dados são liberados com uma licença de uso de dados clara e acessível;
- R1.2. (meta)dados estão associados à proveniência detalhada;
- R1.3. (meta)dados atendem aos padrões da comunidade relevantes para o domínio (Wilkinson *et al.*, 2016, p. 8).

O Quadro 4 caracteriza com maior profundidade os Princípios FAIR, a saber:

**Quadro 4** - Princípios de dados FAIR.

Princípio	Descrição
- F - <i>Findability</i>	Propõe que tanto pessoas quanto computadores precisam encontrar os dados. Metadados atribuídos precisam ser reconhecíveis por

(Encontrabilidade)	máquinas para que os dados sejam encontrados e disponibilizados. Para este fim, os metadados atribuídos necessitam ter um único identificador (PID, <i>Persistent Identifier</i> ); além disso, 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.
- A - <i>Accessibility</i> (Acessibilidade)	Propõe que as autorizações de acesso e uso dos dados devem estar explicitadas, sem que seja necessária a utilização de comunicação ou ferramentas específicas; 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 <i>et al.</i> , 2019, p. 178).
- I - <i>Interoperability</i> (Interoperabilidade)	Propõe que os dados necessitam ser identificáveis para os computadores, de forma que possam ser associados com outros de forma automática. Deve 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.
- R - <i>Reusability</i> (Reusabilidade)	Propõe o fornecimento de metadados associados 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.

Fonte: elaboração própria (2024).

Aventurier (2017) destaca que os Princípios FAIR oferecem diretrizes e práticas amplamente aceitas pela comunidade científica, permitindo que produtores e usuários – sejam humanos ou máquinas – utilizem e citem dados de maneira mais eficiente e precisa. Esses princípios devem ser entendidos como “[...] um guia para a preparação dos dados de pesquisa para reutilização, com condições claramente descritas tanto para pessoas quanto para máquinas” (Imming, 2018, p. 4).

Desde a sua formulação, diversas iniciativas têm buscado promover a adoção dos princípios FAIR. Em 2014, a *Scientific Electronic Library Online* (SciELO) anunciou que começaria a incorporar os princípios FAIR na gestão de seus dados científicos. Nesse contexto, passou a orientar seus usuários a utilizar a ferramenta

"FAIR-Aware<sup>13</sup>", que facilita o processo de tornar os dados depositados na plataforma "o mais FAIR possível" (SciELO, 2021).

Por sua vez, a *GO FAIR Initiative*<sup>14</sup> tem criado um ecossistema aberto e inclusivo para indivíduos, instituições e organizações que colaboram por meio de Redes de Implementação. Essas redes se organizam em três pilares principais:

- o *GO CHANGE*, focado em prioridades, políticas e incentivos para a implementação dos princípios FAIR;
- o *GO TRAIN*, voltado para a coordenação de treinamentos, conscientização e desenvolvimento de habilidades FAIR; e
- o *GO BUILD*, que coordena o desenvolvimento e a implementação das tecnologias necessárias para viabilizar o FAIR.

O debate sobre a abertura de dados de pesquisa tem evoluído e se expandido para incluir outros conceitos relacionados. Um exemplo disso é o conceito de "Dados Vinculados" ou "*Linked Data*", que se refere a um conjunto de princípios para publicar e conectar dados estruturados na *Web*. De acordo com Berners-Lee (2006), dados vinculados são aqueles que estão conectados de forma que possam ser explorados por pessoas ou máquinas, com o objetivo de encontrar outros dados relacionados. Em 2010, Berners-Lee propôs um sistema para avaliar a qualidade da vinculação desses dados, baseado em uma classificação de cinco estrelas, apresentado a seguir:



Disponível na Internet (em qualquer formato; por exemplo, PDF), com licença aberta;



Disponível na Internet de maneira estruturada (em um arquivo Excel com extensão XLS);



Disponível na Internet, de maneira estruturada e em formato não proprietário (CSV em vez de Excel);



Seguindo todas as regras anteriores e dentro dos padrões estabelecidos pelo W3C (RDF e SPARQL): usando URL para identificar coisas e propriedades, de forma que as pessoas possam direcionar para suas publicações;

<sup>13</sup> Disponível em: <https://fairaware.dans.knaw.nl/>.

<sup>14</sup> Disponível em: <https://www.go-fair.org/>.



Seguindo todas as regras anteriores e conectando dados a outros dados, de forma a fornecer um contexto (Isotani; Pinto, 2015, p. 35–36).

Silva (2018, p. 118) observa que o “[...] desenvolvimento de dados abertos interligados ainda é incipiente, mas apresenta um grande potencial de cooperação e reutilização de informações entre pesquisadores com objetivos de estudo comuns”. Nesse contexto, Zátare *et al.* (2019) argumentam que as propriedades dos *Linked Data* devem ser aplicadas em repositórios digitais devido à sua capacidade de integrar dados, promover semântica, aumentar a visibilidade e melhorar a expressividade da linguagem, o que facilita a recuperação de documentos por meio de palavras-chave ou *strings*.

Essas considerações abrem caminho para a discussão sobre as infraestruturas tecnológicas necessárias, surgidas a partir da preocupação dos pesquisadores com a gestão de dados de pesquisa, incluindo *onde* armazenar, disponibilizar e acessar as produções científicas e os objetos informacionais derivados dessas investigações.

## 2.3 REPOSITÓRIOS DIGITAIS

Para o IBICT (2012), um *repositório digital* é uma base de dados *online* que organiza e centraliza a produção científica de uma instituição ou área temática. Esses repositórios desempenham um papel crucial na preservação e na livre circulação da informação científica entre diferentes áreas do conhecimento, sendo fundamentais para o avanço da Ciência. Eles são produtos das Humanidades Digitais, um campo científico interdisciplinar que tem contribuído para o desenvolvimento de soluções que atendem às necessidades informacionais de indivíduos e instituições.

Antes da criação dos repositórios digitais, as produções científicas eram armazenadas em bibliotecas físicas das instituições de origem, com acesso restrito

e circulação limitada. Esse modelo dificultava o acesso e a disseminação do conhecimento, retardando, assim, o avanço das investigações científicas.

O primeiro repositório digital documentado, chamado “ArXiv<sup>15</sup>”, foi criado no início da década de 1990 nos Estados Unidos. Ele armazenava documentos das áreas de Ciência da Computação, Física, Matemática e Ciências Não Lineares. Desde então, com o avanço da tecnologia e a crescente adesão à Ciência Aberta, o volume de informação gerado atingiu proporções tão grandes que seu gerenciamento muitas vezes só é viável com o uso de *softwares* especializados.

Nesse contexto, os repositórios digitais se tornaram sistemas essenciais para o armazenamento e a recuperação de informações, atendendo às novas necessidades de gestão do conhecimento em acesso aberto (Tartarotti; Dal’Evedove; Fujita, 2020).

Originalmente, os repositórios digitais se dividiam em institucionais e temáticos (ou disciplinares), sendo geralmente dedicados ao arquivamento de monografias, como Teses e Dissertações.

Nos *repositórios institucionais*, o foco está em oferecer à comunidade acadêmica um conjunto de serviços para a gestão e divulgação de materiais digitais produzidos pelos membros da instituição. Esses repositórios, que centralizam a produção científica da instituição, visam organizar e disseminar esse conhecimento, conferindo maior visibilidade à sua propriedade intelectual. São fundamentais para a inserção da instituição no ecossistema científico e acadêmico em que está inserida.

De acordo com Kuramoto (2006, p. 142), um repositório institucional compromete-se a “[...] promover e assegurar o controle bibliográfico e facilitar o acesso à informação ali depositada, além de ser responsável pela sua preservação”.

Já os *repositórios temáticos* ou *disciplinares* reúnem a produção científica de uma área específica do conhecimento, sem se limitar às fronteiras institucionais. São especializados e agregam obras relacionadas a um determinado tema, ramo ou área de estudo.

---

<sup>15</sup> Disponível em: [arXiv.org](http://arXiv.org).

Nos últimos anos, tanto os repositórios institucionais quanto os temáticos evoluíram, originando os *repositórios de dados de pesquisa* – plataformas digitais dedicadas exclusivamente ao armazenamento e à disponibilização de dados de pesquisa. Esses repositórios podem ser *institucionais*, reunindo dados de múltiplas áreas do conhecimento de uma mesma instituição, ou *temáticos*, agregando dados específicos de uma área do conhecimento, de uma instituição específica ou ainda de outros agentes, independentemente dos limites institucionais.

Uma outra possibilidade é que os dados de pesquisa passem a ser incorporados em repositórios institucionais já existentes, originalmente dedicados a outros tipos de produções científicas, como Teses e Dissertações, criando *repositórios digitais mistos*. No entanto, Sayão e Sales (2016) argumentam que a melhor prática é depositar dados em repositórios dedicados exclusivamente a eles, pois esses sistemas oferecem uma infraestrutura tecnológica mais adequada para o gerenciamento e a contextualização desses recursos.

Com uma infraestrutura adequada, os repositórios de dados de pesquisa proporcionam suporte essencial aos pesquisadores para a preservação, gestão e disponibilização dos dados, com o objetivo de facilitar seu reuso pela comunidade científica (Monteiro *et al.*, 2017). O crescente uso de repositórios de dados de pesquisa acompanha a formação crescente de novos pesquisadores e a massificação da produção científica em todo o mundo. A subseção seguinte se dedica a discutir mais detalhadamente esse tipo notável e relativamente novo de sistema para organização e recuperação da informação.

### 2.3.1 Repositórios de dados de pesquisa

De acordo com Rice e Southall (2016), os *repositórios de dados de pesquisa* são plataformas que centralizam todos os dados gerados durante a execução de um projeto de pesquisa, disponibilizando-os para reutilização por uma comunidade que vai além da simples busca por informações, exigindo também um envolvimento mais profundo com o processo investigativo. Esses repositórios desempenham um papel

fundamental no fluxo informacional, atendendo às necessidades de informação de pesquisadores, instituições e sociedades científicas, ao mesmo tempo em que aumentam a visibilidade dos resultados das pesquisas e promovem a preservação da memória científica.

Pampel *et al.* (2013) propuseram uma classificação dos repositórios de dados de pesquisa em quatro tipologias principais: 1) *repositório institucional de dados de pesquisa*; 2) *repositório disciplinar de dados de pesquisa*; 3) *repositório multidisciplinar de dados de pesquisa*; e 4) *repositório de dados de pesquisa orientados por projetos*. Essas tipologias estão detalhadas no Quadro 5, conforme segue:

**Quadro 5** - Tipologias de repositórios de dados de pesquisa.

Tipologias	Descrição
Repositório institucional de dados de pesquisa	Funciona no âmbito de uma instituição acadêmica e sendo por ela gerenciada, como universidades ou institutos de pesquisa. São voltados para arquivar dados que são, geralmente, provenientes unicamente das atividades acadêmicas dessas instituições.
Repositório disciplinar de dados de pesquisa	Funciona no âmbito de domínios específicos de pesquisa. Exemplos: Física de Partículas, Ciências Ambientais.
Repositório multidisciplinar de dados de pesquisa	Reúne coleções de dados coletados ou gerados por atividades de pesquisa em várias áreas de conhecimento.
Repositório de dados de pesquisa orientados por projetos	Reúne dados e/ou coleções de dados resultantes de projetos de pesquisa ou resolução de problemas específicos.

Fonte: adaptado de Pampel *et al.* (2013).

Os dados de pesquisa estão sujeitos a diversas ameaças devido às fragilidades inerentes a esses próprios recursos digitais, o que torna essencial garantir sua autenticidade ao longo de todo o seu ciclo de vida. Essa questão representa um grande desafio tanto para as organizações que geram esses dados quanto para as instituições responsáveis pela preservação digital. De acordo com o Conselho Nacional de Arquivos (CONARQ, 2022, p. 7), “o arquivamento e a preservação digital constituem uma questão complexa que envolve muitas variáveis,

compromissos de longo prazo e a necessidade de expressivos investimentos em infraestrutura tecnológica, pesquisa e recursos humanos”.

O CONARQ, órgão vinculado ao Arquivo Nacional do Ministério da Justiça e Segurança Pública, que define a política nacional de arquivos, fundamenta em suas “Diretrizes para Implementação de Repositórios Arquivísticos Digitais Confiáveis<sup>16</sup>” (RDC-ARQ) que os arquivos devem contar com *repositórios digitais confiáveis* para a gestão, preservação e acesso de documentos digitais (CONARQ, 2022). Essas diretrizes apresentam parâmetros para a criação de repositórios que garantam a autenticidade, a preservação e o acesso a materiais digitais durante o tempo necessário. Embora originadas do contexto arquivístico, esses parâmetros são aplicáveis a qualquer repositório digital, incluindo aqueles especializados em dados de pesquisa.

No documento supracitado, o CONARQ estabelece requisitos necessários para um repositório digital confiável subdividindo-os em três conjuntos, sendo eles: *Infraestrutura organizacional; Gerenciamento do documento digital; e Tecnologia, infraestrutura técnica e segurança.*

Quanto à Infraestrutura organizacional, o ambiente em que o repositório digital irá se estabelecer deve cumprir requisitos relativos: à *Governança e viabilidade organizacional; à Estrutura organizacional e de pessoal; à Transparência de procedimentos e arcabouço político; à Sustentabilidade financeira; e à Contratos, licenças e passivos.* Quanto ao Gerenciamento do documento digital, deve cumprir requisitos relativos: à *captura de documentos digitais e à criação do pacote de arquivamento (na admissão); ao Planejamento da preservação; ao Armazenamento e preservação/manutenção do AIP (archival information package); ao Gerenciamento de informação; e ao Gerenciamento de acesso.* Por fim, quanto à Tecnologia, infraestrutura técnica e segurança, deve cumprir requisitos relativos: à *Infraestrutura de sistema; e, à Gestão de risco e segurança* (CONARQ, 2022).

---

<sup>16</sup> Disponível em:

[https://www.gov.br/conarq/pt-br/centrais-de-conteudo/publicacoes/conarq\\_diretrizes\\_rdc\\_arq\\_resolucao\\_43.pdf](https://www.gov.br/conarq/pt-br/centrais-de-conteudo/publicacoes/conarq_diretrizes_rdc_arq_resolucao_43.pdf).

O CONARQ também apresenta uma série de documentos de referência, padrões e normas para a construção de repositórios digitais confiáveis. Entre esses, destacam-se o modelo *Open Archival Information System* (OAIS), o “Modelo de empacotamento de dados”, a “Certificação e auditoria de repositórios confiáveis” (TRAC), os “Metadados de preservação PREMIS”, e o “Padrão de codificação e transmissão de metadados METS”, entre outros (CONARQ, 2022).

Parte significativa do processo de *curadoria de dados digitais*<sup>17</sup> ocorre nos repositórios digitais, onde esses recursos são armazenados. Considerando a relevância para a localização, acesso e reuso dos dados depositados, esta pesquisa foca na *atribuição de metadados*. Os *metadados* são frequentemente definidos como dados que descrevem outros dados, facilitando sua compreensão, acesso e recuperação, além de evitar a perda e a dispersão dos mesmos.

Seguindo as tendências emergentes, diversas instituições têm expandido seus repositórios para incluir dados de pesquisa. No entanto, Sayão e Sales (2016) defendem que a melhor abordagem é arquivar dados em repositórios dedicados exclusivamente a dados, já que esses ambientes tendem a oferecer uma infraestrutura tecnológica mais adequada para contextualização, principalmente por meio da atribuição de metadados. Isso só é viável se os repositórios operarem com infraestruturas confiáveis e estáveis, que garantam a máxima disponibilidade dos serviços.

A perda de dados devido à baixa qualidade da infraestrutura pode ter impactos econômicos e pessoais significativos, forçando a repetição de

---

<sup>17</sup> No âmbito científico, *serviços de curadoria de dados de pesquisa* abarcam ações antes e durante a permanência dos dados em repositórios digitais. Bibliotecários apresentam importante papel nesse contexto, onde, possivelmente, conduzem as seguintes atividades: 1. recebimento da solicitação de depósito dos dados de pesquisa; 2. validação dos dados do pesquisador; 3. recebimento e transferência do Plano de Gestão de Dados do pesquisador para o repositório; 4. recebimento e transferência dos dados de pesquisa do pesquisador para o repositório; 5. conversão dos dados para diferentes formatos e/ou renomeação dos arquivos; 6. *identificação, adicionando metadados e descritores aos conjuntos de dados*; 7. determinação do período de preservação do conjunto de dados (curto, longo ou permanente); 8. determinação do nível de acesso aos dados (público, embargado, restrito); 9. carregamento e publicação do Programa de Gestão dos Dados e dos dados depositados; 10. atribuição do *Digital Object Identifier* (DOI) aos dados e a sua ativação; 11. garantia da descoberta desses dados pelos mecanismos de busca como *Google Data*, por exemplo; 12. mantimento dos dados atualizados replicados em outros repositórios, para garantir sua preservação e acesso; 13. análise do acesso e reuso dos dados; e, 14. elaboração de Guias e Tutoriais baseados nas Políticas Institucionais e melhores práticas.

experimentos científicos, o que pode resultar em oportunidades perdidas para sempre (Lynch, 2008).

De acordo com Viana (2020, p. 49), a compreensão dos dados de pesquisa “[...] depende do contexto em que foram gerados, sendo necessário que os pesquisadores forneçam informações contextuais por meio de metadados ou outros instrumentos.” Assim, os metadados são essenciais para a contextualização dos dados de pesquisa e contribuem para a organização da informação.

A NISO (2004, p. 1) define metadados como “[...] a informação estruturada que descreve, explica, localiza ou facilita a recuperação, o uso ou a gestão de um recurso informacional”. No contexto de dados digitais, os metadados devem fornecer descrições suficientes para possibilitar o entendimento completo do objeto informacional. Ressalta-se que, esses recursos apresentam complexidade e desafios em sua organização, o que exige métodos apropriados para sua gestão. Duas condições principais são essenciais para garantir o sucesso nesse processo: a primeira é uma *infraestrutura tecnológica e gerencial robusta*, que permita o acompanhamento dos dados durante todo o seu ciclo de vida de forma sustentável; a segunda é o *fornecimento adequado de metadados descritivos e temáticos* (Sayão; Sales, 2016).

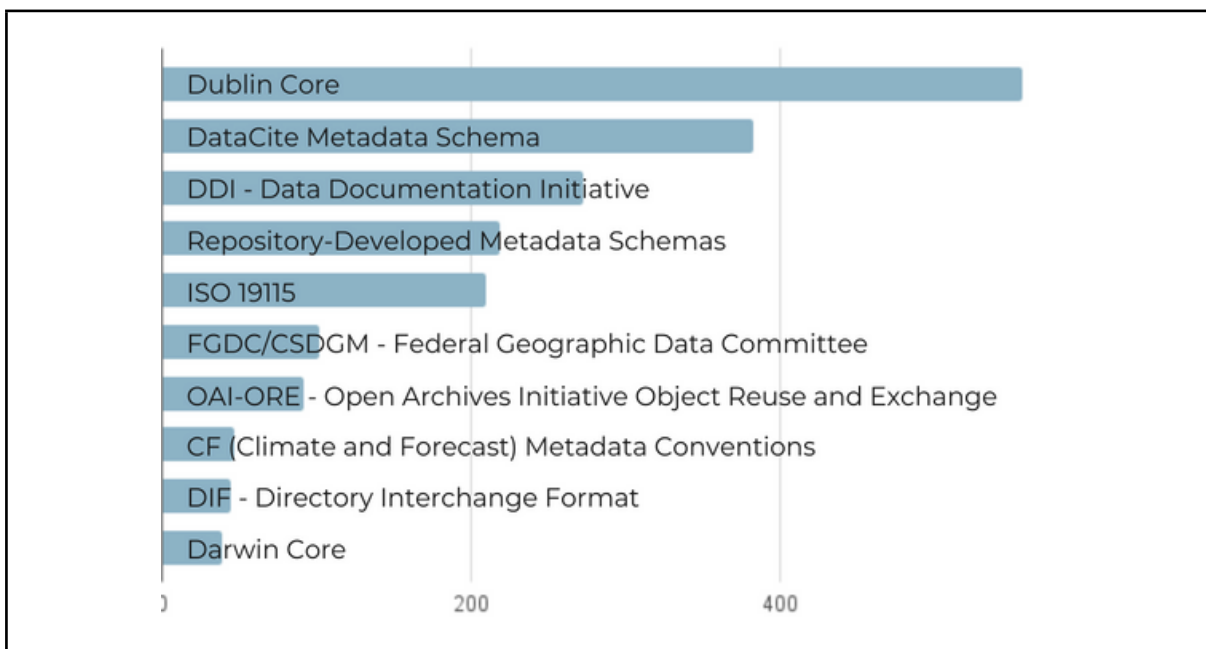
Metadados descrevem os atributos de um objeto informacional digital, como sua natureza, estado, estrutura e semântica. Hodson *et al.* (2018) destacam que os dados devem ser acompanhados por uma pluralidade de atributos relevantes. No caso dos dados de pesquisa, os metadados oferecem significado semântico a dados isolados, proporcionando o contexto necessário para facilitar a busca e a recuperação da informação (Márdero-Arellano; Shintaku, 2005, p. 15). Sayão e Sales (2016, p. 99) ressaltam que, devido à heterogeneidade dos dados de pesquisa, os repositórios digitais devem ser dinâmicos e flexíveis, a fim de apoiar adequadamente os processos de gestão desses dados.

Nos repositórios digitais, os metadados são organizados em esquemas, que normalizam e descrevem as informações, criando padrões que facilitam a busca, o acesso e a recuperação dos objetos armazenados. Esses esquemas possuem

componentes semânticos e de conteúdo, e são codificados com sintaxes ou linguagens de marcação<sup>18</sup>. O uso de esquemas específicos para um ambiente informacional facilita significativamente o processo de descoberta de dados.

Existem vários esquemas de metadados, sendo que a plataforma re3data<sup>19</sup> documenta os padrões mais utilizados por repositórios de dados de pesquisa registrados nela. A Figura 1 ilustra esses padrões.

**Figura 1** - Padrões de metadados mais utilizados em repositórios de dados de pesquisa.



Fonte: re3data (2023).

Cada esquema de metadados possui características específicas, sendo mais adequado para descrever determinados tipos de documentos e atender a necessidades de comunidades distintas. Alguns padrões de metadados também são amplamente adotados em áreas científicas específicas. Gularte (2021), com base em dados do re3data, identificou os dez padrões de metadados mais utilizados em 2021, destacando as áreas científicas a que estão mais estreitamente relacionados, conforme ilustrado no Quadro 6.

<sup>18</sup> *Standard Generalized Markup Language (SGML)* e *Extensible Markup Language (XML)* são exemplos de linguagem de marcação.

<sup>19</sup> *Registry of Research Data Repositories*. Disponível em: <https://www.re3data.org/>.

**Quadro 6** - Padrões de metadados por área do conhecimento.

Padrão de metadado	Área do Conhecimento
Dublin Core	Geral
DataCite	Geral
Data Document Initiative	Ciências Sociais
Repository Developed Metadata Schemas	Geral
ISO 19115	Ciências Espaciais
Content Standard for Digital Geospatial Metadata (CSDGM)	Ciências Espaciais
Directory Interchange Format (DIF)	Ciências Espaciais
Climate and Forecast (CF) Metadata Conventions	Ciências Espaciais
Ecological Metadata Language (EML)	Biologia
Open Archives Initiative Object Reuse and Exchange (OAI-ORE)	Geral

Fonte: Adaptado de Gularte (2021).

O *Dublin Core* é um esquema de metadados amplamente utilizado para a descrição de recursos eletrônicos. Desenvolvido como uma alternativa para simplificar e otimizar a descrição e recuperação de objetos informacionais digitais na *web*, o *Dublin Core* segue um padrão mínimo que garante a interoperabilidade entre sistemas. Adotando a sintaxe do *Resource Description Framework* (RDF), uma estrutura geral para representar dados interconectados na *web*, o *Dublin Core* oferece uma semântica universalmente compreensível entre os metadados.

Nos repositórios digitais, é fundamental garantir que todos os metadados necessários para descrever adequadamente os dados sejam contemplados pelo esquema escolhido. Muitas vezes, a combinação de diferentes esquemas de metadados pode ser a solução mais eficaz. Segundo Terra *et al.* (2021), esquemas como o *Dublin Core* podem ser adaptados ao contexto de cada instituição, permitindo a interoperabilidade e o compartilhamento de registros, o que aumenta a visibilidade de produções e recursos científicos.

O desenvolvimento de repositórios digitais voltados para dados de pesquisa é uma tarefa complexa, que exige um diálogo aprofundado entre cientistas da computação, especialistas em dados, profissionais da informação e os próprios

usuários dos sistemas, além das diversas áreas do conhecimento envolvidas. Sendra, Late e Kumpulainen (2023) destacam que há uma grande heterogeneidade, tanto em relação aos dados quanto aos grupos de usuários desses repositórios, o que representa um desafio significativo para a padronização dos sistemas. Serviços de informação podem se deparar com incertezas, refletidas em suas políticas, que afetam a percepção dos depositantes sobre o sistema. Nesse contexto, fornecer documentação clara e transparente torna-se um aspecto essencial de apoio, contribuindo para a adoção e utilização eficiente dos repositórios.

Ribes (2014) previu que os próximos projetos de infraestrutura de pesquisa e ciberinfraestruturas enfrentariam desafios únicos, distintos daqueles encontrados nos últimos trinta anos. Para superá-los de maneira eficaz, foi (e ainda é) necessário adotar uma abordagem sociotécnica no desenvolvimento de repositórios, que permita entender melhor as necessidades das instituições e dos depositantes. Golub *et al.* (2020) alertam que não envolver os grupos de usuários finais nas fases iniciais de criação de infraestruturas de pesquisa pode ser um erro crítico para o sucesso dos serviços de informação. Sendra, Late e Kumpulainen (2023) afirmam que, nas infraestruturas digitais voltadas para dados, os recursos precisam ser tão flexíveis e interoperáveis quanto possível, a fim de atender às necessidades de todos os tipos de usuários finais.

Para concluir a discussão sobre alguns princípios, elementos e práticas da Ciência Aberta relevantes para o escopo desta pesquisa, a próxima subseção abordará a prática do autoarquivamento de produções e recursos científicos em repositórios digitais.

### *2.3.1.1 A prática do autoarquivamento*

Em muitos sistemas de recuperação da informação, os próprios autores podem inserir suas produções e recursos científicos de maneira direta, sem a intermediação de profissionais especializados – prática conhecida como *autoarquivamento*. A *Budapest Open Access Initiative* (BOAI, 2002) aponta o

autoarquivamento como uma estratégia crucial para os repositórios digitais baseados no Acesso Aberto, sendo particularmente reforçada com a implementação da Via Verde, em 2002.

Veiga e Macena (2015) explicam que o autoarquivamento permite ao autor tornar sua produção científica acessível a qualquer usuário com acesso à *internet*, o que aumenta significativamente sua visibilidade. Essa prática é incentivada, entre outros motivos, por facilitar a rápida divulgação dos resultados da pesquisa. No entanto, Xia (2012) observou que muitos autores se mostram desmotivados em adotar o autoarquivamento. Keefer (2007) identificou várias razões para essa resistência, incluindo a falta de compreensão sobre a importância da prática no contexto do Acesso Aberto, a falta de tempo para executá-la, e o receio de perder o controle sobre sua obra, incluindo o risco de plágio.

Outro fator que pode agravar essa resistência é a complexidade dos manuais de autoarquivamento, a exemplo do “Manual de Autodepósito de Dados de Pesquisa no RI UFSCAR”, que frequentemente são excessivamente detalhados e com muitas etapas, tornando sua leitura cansativa e, conseqüentemente, dificultando sua adoção.

Suber (2004) argumenta que um dos maiores obstáculos ao avanço do Acesso Aberto na Ciência é a inércia dos próprios autores. Para superar essa dificuldade na adoção da cultura do autoarquivamento, diversas instituições, especialmente aquelas que recebem financiamento público para pesquisas, implementaram políticas obrigatórias. A Universidade de Southampton, no Reino Unido, foi pioneira ao adotar uma política de autoarquivamento obrigatório em 2003. Desde então, a prática se tornou cada vez mais comum. Leite (2009, p. 1) reforça que "os repositórios com mais sucesso em relação ao seu povoamento são aqueles cujas instituições implementaram políticas de depósito obrigatório", sendo este um fator fundamental para aumentar significativamente as taxas de depósito.

A literatura científica sobre o autoarquivamento destaca que essa prática traz benefícios substanciais para todos os envolvidos em ambientes informacionais digitais (Medeiros, 2010; Tartarotti; Dal'Evedove; Fujita, 2019; Freitas, 2019; Silva,

2022). Contudo, no que tange à representação temática da informação, há lacunas que precisam ser abordadas para garantir que os objetos informacionais autoarquivados possam ser acessados de maneira eficaz.

No processo de autoarquivamento, os próprios autores agem como indexadores, embora muitas vezes sem o conhecimento técnico necessário para atribuir descritores e palavras-chave de forma precisa. Para que esses autores possam desempenhar esse papel de maneira eficaz, é essencial que os sistemas de repositórios forneçam diretrizes claras e compreensíveis sobre como realizar a indexação, tornando esse processo uma das responsabilidades-chave desses sistemas.

Os autores devem estar cientes de que a atribuição inadequada de metadados pode prejudicar a futura acessibilidade e visibilidade de suas produções. A escolha assertiva de metadados, como palavras-chave, pode determinar se o material será facilmente encontrado por meio de estratégias de busca, o que impacta diretamente no seu reconhecimento e utilização dentro da comunidade científica. Por outro lado, a atribuição inadequada pode tornar o material "silenciado", dificultando seu acesso e, conseqüentemente, a sua citação e valorização (Brandau; Monteiro; Braile, 2005).

Tartarotti, Dal'Evedove e Fujita (2020, p. 17) destacam que a falha na recuperação de materiais em sistemas de informação "compromete a relevância, visibilidade, *status* e valor público dos repositórios digitais, impactando os indicadores da qualidade científica, social e econômica das instituições". Para mitigar as implicações negativas dessa prática, os repositórios digitais podem adotar *vocabulários controlados* e realizar a *validação dos dados preenchidos pelos autores*, a fim de garantir a qualidade dos metadados temáticos inseridos. No entanto, pesquisas recentes indicam que essas práticas ainda não são amplamente utilizadas (Freitas; Dal'Evedove; Tartarotti, 2021; Silva, 2022), o que reforça a importância de estabelecer critérios claros para a atribuição de metadados.

Diante disso, a próxima subseção irá explorar o universo conceitual e teórico da representação temática da informação, com foco específico na *indexação*. Em repositórios digitais, o conhecimento dos autores sobre os processos de indexação é

crucial para garantir a recuperação eficiente e o reuso adequado dos dados de pesquisa, otimizando o acesso e a disseminação do conhecimento científico.

## 2.4 REPRESENTAÇÃO TEMÁTICA DA INFORMAÇÃO E INDEXAÇÃO

Kobashi (1996), ao discutir o conceito de "representação" no âmbito da Documentação, define-o como

[...] um conceito pré-teórico, associado, de um lado, à descrição de aspectos que identifiquem materialmente os documentos (catalogação) e, de outro, ao processo e ao produto da condensação de conteúdos de textos, ou seja, à indexação e à elaboração de resumos (processos) e aos próprios índices e resumos (produtos).

Para Gomes e Lima (2021, p. 2), a representação da informação não pode ser dissociada de sua recuperação, pois o principal objetivo da representação é fornecer pontos de acesso aos objetos informacionais, facilitando sua identificação e recuperação em um conjunto de documentos.

Em termos gerais, a catalogação lida essencialmente com os aspectos físicos dos documentos, sejam eles organizados em acervos físicos ou digitais. Contudo, o foco deste trabalho recai sobre a *representação temática da informação*<sup>20</sup>.

Embora a indexação seja fundamental para a representação por assunto e a localização de objetos informacionais físicos em sistemas de armazenamento e recuperação da informação, sua relevância tem sido cada vez mais reconhecida no contexto digital. Na *internet*, a indexação permite que o usuário recupere objetos informacionais digitais por meio de buscas baseadas em algum aspecto do seu conteúdo. Nesse contexto, ressalta-se a complexidade pragmática advinda do surgimento de novos tipos de recursos a, inelutavelmente, terem seu conteúdo informacional representado. Para além da indexação dos convencionais livros e monografias, na atualidade, fazem-se necessárias discussões mais acaloradas relativas à indexação, por exemplo, dos, já abordados, dados de pesquisa; sendo

---

<sup>20</sup> Compreende *indexação e resumos*.

esses, recursos que, por sua natureza, demandam orientações específicas para a sua representação temática.

A Associação Brasileira de Normas Técnicas (ABNT, 1992, p. 2) define *indexação* como “o ato de identificar e descrever o conteúdo de um documento com termos representativos de seus assuntos, os quais constituem uma linguagem de indexação”. A *linguagem de indexação*, também chamada de *linguagem documentária*, é uma ferramenta para controle terminológico dentro de um domínio do conhecimento, apresentando termos preferidos que, quando utilizados de forma coordenada na representação e recuperação da informação, possibilitam a comunicação eficiente entre o produtor e o consumidor da informação.

Lancaster (2004) divide o processo de indexação em duas etapas principais e consecutivas: a “análise conceitual” e a “tradução”.

Na primeira etapa, a *análise conceitual*, o indexador deve determinar o conteúdo do documento e identificar seu assunto. Essa etapa deve responder a três perguntas centrais: “Sobre o que trata o documento?”, “Por que ele foi incorporado ao acervo?” e “Quais aspectos do documento serão de interesse para potenciais usuários?”.

Adotando a abordagem de “indexação centrada no usuário”, Lancaster (2004, p. 9) destaca que “os indexadores precisam saber mais do que os princípios da indexação; é essencial que estejam profundamente familiarizados com os interesses da comunidade atendida e as necessidades de informação de seus membros”. Portanto, uma indexação eficaz não se limita a identificar o que o documento trata, mas também a entender por que ele é relevante para um grupo específico de usuários.

A segunda etapa, a *tradução*, consiste em converter a análise conceitual do documento em um conjunto de termos de indexação. Esse processo pode ocorrer de duas maneiras: *por extração*, quando palavras ou expressões<sup>21</sup> do próprio documento

---

<sup>21</sup> Palavras e expressões como as contidas em títulos, resumos, sumários, ilustrações, gráficos, tabelas (Lancaster, 2004).

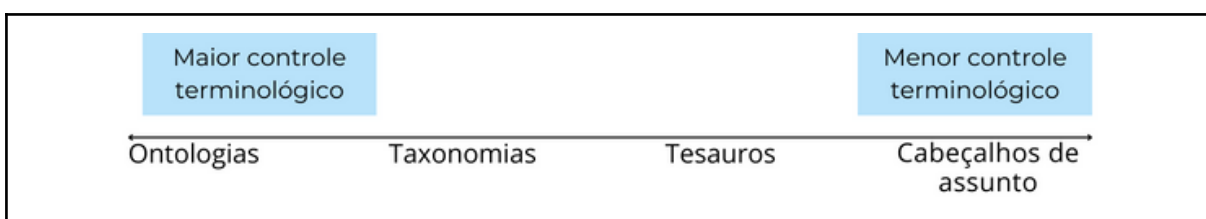
são utilizadas na indexação; ou *por atribuição*, quando os termos são retirados de fontes externas ao documento, como vocabulários controlados.

*Vocabulários controlados*, oriundos de Sistemas de Organização do Conhecimento (SOC), são conjuntos de termos que controlam sinônimos, distinguem homógrafos e agrupam termos afins, criando uma terminologia contextualizada em um domínio específico do conhecimento. O objetivo do controle de vocabulário é garantir a consistência terminológica na indexação, evitando a dispersão de objetos informacionais relacionados a um termo descritor específico.

Em sistemas de informação científica, a terminologia utilizada na indexação deve ser altamente especializada e precisa, pois a linguagem científica está diretamente associada ao vocabulário técnico e acadêmico. Como Lancaster (2004, p. 10) argumenta: "Quanto mais especializada a clientela de um centro de informação, maior a probabilidade de que a indexação deva ser feita de forma personalizada, ajustando-se com precisão aos interesses do grupo."

Exemplos de vocabulários controlados incluem Ontologias, Tesouros, Taxonomias e Cabeçalhos de Assunto. A Figura 2, a seguir, classifica esses vocabulários em termos do nível de controle terminológico, desde o mais rigoroso até o menos controlado.

**Figura 2** – Vocabulários controlados e níveis de controle de vocabulário.



Fonte: elaboração própria (2024).

Lancaster (2004) argumenta que o tipo de vocabulário controlado não é o fator mais importante na etapa de tradução da indexação; o que realmente influencia a qualidade do processo são o alcance (abrangência) e a especificidade do vocabulário utilizado. Em especial, ele destaca a importância de se considerar as

combinações de termos de indexação, e não apenas os termos isolados (Lancaster, 2004).

Os termos utilizados na indexação funcionam como substitutos do próprio documento, permitindo que, ao buscar uma informação, o usuário não precise acessar o documento na íntegra. Além disso, esses termos devem ter um alto grau de representatividade do conteúdo do documento, garantindo uma recuperação eficiente e precisa. Nesse sentido, Strehl (1998, p. 1) afirma que “o que determina a qualidade da indexação é o estabelecimento de descritores que representem devidamente os assuntos dos documentos, permitindo sua eficaz recuperação”.

Embora o processo de indexação seja guiado pelas duas etapas propostas por Lancaster (análise conceitual e tradução), ele ressalta que a indexação não segue regras absolutas, pois envolve relativa flexibilidade e subjetividade. Pinheiro (1978, p. 109) identifica a indexação como um processo crítico que depende do julgamento individual do indexador, o que pode gerar variações nas escolhas de termos e descritores. No entendimento da referida autora, diferentes indexadores, ou um mesmo indexador em momento distintos, pode perceber de forma diferente desde os conceitos mais adequados para representarem o conteúdo de um documento, até os descritores mais adequados para representar esses conceitos (Pinheiro, 1978). Em concordância, Fujita e Santos (2016) apontam que as competências do indexador influenciam diretamente a qualidade da indexação e, conseqüentemente, a eficácia na recuperação da informação.

Com o avanço dos sistemas de recuperação de informação na *internet*, novas práticas de indexação surgiram. Ao contrário das bibliotecas físicas, onde a atribuição de metadados era realizada exclusivamente por bibliotecários ou profissionais especializados, na *internet* é comum que os próprios usuários, como os autores dos documentos, preencham os metadados e escolham os termos descritores e palavras-chave.

Com o objetivo de alcance de uma maior coerência na indexação praticada (*inter indexadores e intra-indexador*<sup>22</sup>), uma maior fidedignidade temática e consistência terminológica, também podem ser indicados a esses usuários a utilização de vocabulários controlados para que embase a definição dos termos. Outra prática cabível é a validação de palavras-chave e termos descritores por profissionais habilitados do sistema.

De acordo com Fattahi (1998), a uniformidade e a consistência de termos de acesso a documentos digitais, potencializados pelo uso de instrumentos de indexação, são requisitos básicos para o efetivo controle e acesso no ambiente global *online*.

A *análise de assunto*, intrínseco ao processo de indexação, deve ser orientada por princípios vinculados expressos, preferivelmente, em uma Política de Indexação. Esses princípios, sendo eles, o de *especificidade*, de *exaustividade*, de *revocação* e de *precisão*, devem influenciar o indexador na sua decisão sobre a determinação de conceitos e termos descritores e o resultado poderá ser observado pelo usuário na recuperação da informação.

A *especificidade* está relacionada ao nível de abrangência que um sistema de informação e que um vocabulário controlado permitem especificar os conceitos identificados em um documento. Por exemplo, um documento cujo assunto seja especificamente "pinguins", ao ser indexado com o assunto "aves", refletirá uma baixa especificidade na definição de assuntos, que por sua vez, ocasionará uma alta capacidade de revocação na recuperação.

A capacidade de *revocação* diz respeito ao número de documentos recuperados mediante uma busca em um sistema de recuperação de informação e pode ser mensurada através da relação entre o número de documentos relevantes sobre determinado tema recuperados pelo sistema e o número total de documentos sobre o assunto, existentes nos registros do mesmo sistema.

---

<sup>22</sup> Strehl (1998, p. 1) define que a coerência da indexação "[...] consiste na concordância quanto aos termos usados para indexar um assunto tratado em diferentes documentos". De acordo com Lancaster (2004, p. 68), "a coerência inter indexadores refere-se à concordância entre indexadores, enquanto a coerência intra-indexador refere-se à extensão com que um indexador é coerente consigo mesmo".

A capacidade de *precisão*, ou *relevância*, está relacionada ao número de documentos recuperados em um sistema que atende às necessidades de informação do usuário. Ela pode ser mensurada por meio da relação entre os documentos relevantes recuperados e número total de documentos recuperados.

A indexação realizada com maior especificidade terminológica resultará, portanto, em uma recuperação com níveis de revocação menores e com maiores índices de precisão. Ou seja, ainda que seja recuperado um número reduzido de documentos, é maior a probabilidade que esses documentos correspondam às necessidades do usuário.

A quantidade de termos descritores utilizados para descrever um documento diz respeito à *exaustividade*, que, por sua vez, está relacionada à revocação e à precisão do sistema de recuperação da informação. Por exemplo, se um documento trata especificamente dos assuntos "pinguim-rei", "pinguim-imperador" e "pinguim-gentoo" e o indexador optar por uma maior exaustividade na indexação, ele poderá considerar na indexação cada um desses termos e, ainda, o termo "aves". Uma busca no sistema efetuada com a submissão do termo "aves" recuperará esse documento, assim como uma busca efetuada com qualquer um dos termos "pinguim-rei", "pinguim-imperador" e "pinguim-gentoo".

Lancaster (2004, p. 4) diferencia revocação e precisão da seguinte forma: "[Revocação] é a capacidade [do sistema] de recuperar documentos úteis, enquanto precisão é a capacidade de evitar documentos inúteis". Desse modo, uma alta revocação está ligada a uma baixa precisão na atribuição de termos; por outro lado, quanto mais específica for a definição de termos, mais precisa se revelará a recuperação e com menor índice de revocação.

A indexação realizada com menor especificidade tende a considerar apenas os assuntos principais do documento, enquanto uma indexação realizada com maior exaustividade, tende a considerar o maior número possível de assuntos, incluindo assuntos principais e secundários. Quanto mais exaustiva for a indexação, mais alta tende ser a revocação, porém, com baixa precisão da recuperação; em contrapartida, quanto mais específica for a seleção de termos, mais precisa será a recuperação, com baixa revocação de documentos (Lancaster, 2004).

O estilo de indexação escolhido, considerando os referidos princípios, trará implicações positivas ou negativas na recuperação da informação e deve depender da estratégia política adotada pela respectiva instituição. Nesta ótica, exaustividade e especificidade “[...] são elementos que devem estar claramente definidos na política do sistema de informação, pois exercerão influência direta sobre a indexação realizada” (Rubi; Fujita, 2003, p. 70).

Linguagens documentárias, quando atualizadas e alinhadas com as filosofias da instituição, amparam-nas de diversas maneiras, como no combate de termos descritores homógrafos, de erros ortográficos e de expressões regionais; na determinação de termos singulares e plurais; nas questões relacionadas à polissemia e sinonímia; e no ajuste da profundidade e especificidade da descrição. Boccato e Torquetti (2012, p. 79) afirmam que “o uso de uma linguagem de indexação controlada permite a representação na indexação e a pesquisa por assunto com maior exatidão e equidade com as necessidades informacionais dos usuários”. Além disso, essas linguagens apresentam termos preferidos/autorizados que facilitam a comunicação de profissionais e usuário mediante a padronização terminológica, controlando dispersões lexicais para a descrição dos conteúdos.

Sistemas de Organização do Conhecimento, em geral, podem ser fundamentais para o esforço científico, desde que haja adaptação às novas temporalidades do desenvolvimento teórico e técnico da respectiva área do conhecimento (Hjørland, 2012; Ibekwe-Sanjuan; Bowker, 2017). Contudo, se por um lado, são perceptíveis vantagens no uso desses sistemas, por outro lado, também são revelados problemas em seu emprego, principalmente quando não acompanham a evolução e o surgimento de termos do respectivo domínio do conhecimento.

No passado, os vocabulários controlados eram desenvolvidos para atender às necessidades de profissionais especializados em indexação e pesquisa. Contudo, atualmente, há uma demanda crescente por vocabulários mais intuitivos, que possam ser usados de maneira eficiente por qualquer usuário de sistemas de recuperação de informação.

Diante dos desafios contemporâneos da indexação, a próxima seção do estudo se dedica à análise da indexação de documentos utilizando *palavras-chave*, tema central desta pesquisa.

## 2.5 PALAVRAS-CHAVE NA REPRESENTAÇÃO TEMÁTICA DA INFORMAÇÃO

Nas discussões sobre Organização do Conhecimento, alguns pesquisadores destacam a necessidade de diferenciar os conceitos de "palavras-chave" e "descritores". Gonçalves (2008, p. 5), por exemplo, aponta que a atribuição de palavras-chave a documentos não segue uma estrutura terminológica rigorosa, sendo, em muitos casos, aleatória. Segundo o autor, em documentos com conteúdo textual, como artigos científicos, palavras-chave podem ser extraídas de qualquer parte do texto, como do título, do resumo e do próprio corpo do documento, sendo atribuídas de forma livre por seus autores, editores ou geradas automaticamente. Brandau, Monteiro e Braile (2005, p. 8) afirmam que, para uma palavra-chave se tornar um descritor, ela deve passar por um controle rigoroso de sinônimos, significados e relevância dentro de uma estrutura de classificação de um determinado assunto.

Palavras-chave são termos representativos do conteúdo de um texto ou objeto informacional, utilizados para facilitar a recuperação da informação. De acordo com Aquino e Aquino (2013), no contexto de documentos textuais, as palavras-chave referem-se às expressões mais significativas de um texto que podem ser utilizadas para indexação. Além de representarem o conteúdo do documento, elas servem também para expressar as necessidades informacionais do usuário durante o processo de busca (Fujita, 2004, p. 257).

Essas palavras-chave funcionam como uma condensação do conteúdo do documento, com um grau de abstração superior ao do título ou resumo. Fujita (2004, p. 257) afirma que as palavras-chave são uma representação significativa do texto, podendo ser vistas como um "resumo encapsulado" de seu conteúdo. Segundo Fujita

e Tartarotti (2020, p. 332), com a chegada da era digital e da *web*, as palavras-chave tornaram-se um elemento essencial nos sistemas de armazenamento e recuperação da informação, sendo fundamentais para garantir que objetos informacionais digitais sejam localizados por mecanismos de busca e atinjam seus usuários-alvo.

No contexto do autoarquivamento, onde os próprios autores são responsáveis pela indexação e encontrabilidade de suas produções, a atribuição de palavras-chave que representem de forma precisa o conteúdo de artigos científicos, dados de pesquisa e outros objetos informacionais, se torna uma prática essencial. Como especialistas nos temas abordados, os autores devem ser capazes de identificar e utilizar as palavras-chave mais adequadas para a representação de seus documentos. Para isso, é necessário um conhecimento profundo do seu campo de pesquisa e domínio científico, bem como uma compreensão clara da organização do conteúdo do documento, para que os termos reflitam adequadamente seus tópicos principais e secundários (Fujita; Tartarotti, 2020).

Para Fujita e Tartarotti (2020, p. 337), a atribuição de palavras-chave a produções científicas, embora aparentemente simples, também deve envolver padronização, equivalência e consistência entre as demais produções do mesmo autor, “[...] tendo em vista que as palavras-chave serão posteriormente adotadas para a definição de perfil científico, em diversas fontes de informação”. Como observa Silva (2022, p. 71), a atribuição de palavras-chave é crucial não apenas para a recuperação da informação, mas também para garantir visibilidade ao documento, ao seu autor e à instituição a que pertencem.

Estratégias específicas de indexação são fundamentais para garantir a precisão na atribuição de palavras-chave. Por exemplo, a ordem de apresentação das palavras-chave pode ser importante em alguns sistemas de recuperação da informação. Em muitos casos, recomenda-se que as palavras-chave associadas ao tema central do documento sejam registradas primeiro, seguidas pelas palavras-chave relacionadas aos tópicos secundários (Silva, 2022). Em sistemas de recuperação da informação que fornecem orientações para a atribuição de palavras-chave, costuma-se disponibilizar campos de preenchimento específicos para essas palavras, ajudando o usuário a seguir boas práticas.

Diversos estudos têm explorado as implicações da atribuição de palavras-chave por autores de artigos científicos. Gonçalves (2008), por exemplo, analisou a atribuição de palavras-chave em artigos das áreas de Antropologia, Ciência Política e Sociologia, e concluiu que, dependendo da estratégia adotada pelos autores, as palavras-chave podem complementar o conteúdo dos resumos e servir como indicadores de relações conceituais. No entanto, o estudo indicou que essa função estratégica das palavras-chave é pouco explorada nas Ciências Sociais.

Em pesquisa realizada na Universidade de Coimbra, Miguéis *et al.* (2013) compararam as palavras-chave atribuídas pelos autores a artigos científicos no MEDLINE com os termos de linguagem controlada presentes no *Medical Subject Headings* (MeSH). Os resultados mostraram que as palavras-chave atribuídas pelos autores frequentemente não correspondiam às utilizadas pela editora da revista e que o número de palavras-chave atribuídas era consideravelmente menor do que o número de termos controlados presentes no MEDLINE.

Um estudo mais recente conduzido por Terra *et al.* (2021), que analisou 48.501 registros de metadados de Dissertações no repositório digital da Universidade de São Paulo (USP), observou discrepâncias entre as palavras-chave atribuídas em português e em inglês. O estudo constatou uma frequência alta de repetição de palavras do título nas palavras-chave, tanto em português (33,01%) quanto em inglês (33,54%). No entanto, foi observada uma baixa taxa de repetição de palavras-chave no resumo (11,27% em português e 11,46% em inglês). Terra *et al.* (2021) argumentaram que, para este caso, a repetição das palavras do título nas palavras-chave não necessariamente melhora a recuperação da informação, e destacaram a importância de incluir palavras-chave no resumo, visto que o resumo consolida o desenvolvimento semântico do conteúdo.

Com base nessas observações, os autores sugerem que algumas práticas poderiam melhorar a qualidade da indexação no sistema de recuperação de informação investigado. Essas práticas incluem: (i) garantir a mesma quantidade de palavras-chave em português e em inglês; (ii) manter a mesma ordem de apresentação das palavras-chave nos dois ou mais idiomas; (iii) oferecer vocabulário controlado em forma de lista suspensa, para facilitar a escolha das palavras-chave;

(iv) fornecer informações sobre a consistência entre as palavras no título, resumo e palavras-chave; (v) disponibilizar diretrizes claras para o processo de autoarquivamento; e (vi) oferecer recursos adicionais, como tutoriais e listas de perguntas frequentes, para orientar os depositantes na atribuição de palavras-chave (Terra *et al.*, 2021).

Esses estudos revelam o panorama negativo da indexação realizada sem o apoio de ferramentas que garantam padronização e consistência. Embora as palavras-chave desempenhem um papel crucial na recuperação da informação, nem todos os sistemas fornecem aos autores científicos as orientações necessárias para a definição adequada desses termos durante o processo de autoarquivamento de objetos informacionais digitais. A próxima subseção irá dissertar sobre *diretrizes* e recursos estratégicos que podem orientar os depositantes na indexação, com o objetivo de melhorar a representatividade temática dos documentos.

## 2.6 DIRETRIZES E POLÍTICAS PARA ATRIBUIÇÃO DE TERMOS DESCRITORES E PALAVRAS-CHAVE EM SISTEMAS DE AUTOARQUIVAMENTO

Diretrizes formalizadas para a atribuição de palavras-chave e termos descritores desempenham um papel fundamental na orientação dos autores científicos durante o processo de autoarquivamento. Com essas orientações à disposição, os pesquisadores podem, de maneira mais racional e objetiva, selecionar os termos mais adequados para os documentos a serem depositados em sistemas digitais de armazenamento. O principal objetivo dessas diretrizes é promover a padronização das operações de indexação, garantindo a coerência e a consistência dos termos escolhidos.

Essas diretrizes frequentemente fazem parte de documentos como "Políticas de Indexação", "Políticas de Autoarquivamento", "Manuais de Autoarquivamento", "Guias de Autoarquivamento" e "Tutoriais para Autoarquivamento". Embora, em teoria, cada um desses documentos tenha características distintas e enfoques variados, todos compartilham o objetivo comum de fornecer critérios – com

diferentes níveis de profundidade – para a atribuição de palavras-chave e termos descritores aos documentos submetidos.

Dedicado, de maneira ampla, a equipes editoriais e ao contexto da curadoria de dados de pesquisa, como exemplo, a Scielo (2023) disponibiliza em seu acervo normativo um documento denominado “Guia de curadoria de dados de pesquisa para equipes editoriais” que registra um conjunto de diretrizes para que equipes editoriais depositem dados de pesquisa na referida biblioteca virtual. Este documento considera, dentre outras, questões relativas ao preenchimento de metadados, indicando, ainda que brevemente, *como se definir e incluir palavras-chave*. A elaboração de diretrizes para a indexação, mesmo quando incluídas em documentos de conteúdo amplo, pode ser destacada como um empreendimento relevante na idealização de um sistema de recuperação da informação, uma vez que impacta no seu desenvolvimento e, por consequência, na percepção dos seus usuários quanto ao seu fim.

Nos últimos anos, estudos nacionais têm evidenciado a relevância de documentos com diretrizes claras para a atribuição de palavras-chave a dados de pesquisa, especialmente no contexto de autoarquivamento. Destacam-se, entre esses estudos, o de Silva (2022), intitulado “Atribuição de palavras-chave na prática do arquivamento: estudo das diretrizes de repositórios de dados de pesquisa ibero-americanos”; e o de Viana e Dal’Evedove (2021), “Indexação de dados científicos: uma análise a partir das políticas da Rede de Repositórios de Dados Científicos do Estado de São Paulo”.

Contudo, a investigação da prática nesses estudos revela problemas recorrentes relacionados à falta de documentos com diretrizes adequadas ou à sua baixa qualidade, além da insuficiente capacitação dos profissionais envolvidos no processo de indexação. Quando não é uma questão de desconhecimento sobre a importância da existência, adequação e disponibilidade de diretrizes bem definidas para a indexação, o problema pode decorrer de uma simples omissão por parte dos administradores dos sistemas, que muitas vezes encaram a indexação apenas como uma tarefa técnica sem a necessidade de um processo sistematizado de identificação de assuntos (Rubi; Fujita, 2006).

Pesquisadores do campo da Ciência da Informação, como Fujita (2021), Oliveira *et al.* (2021), Andrade e Sabbag (2017), defendem que todos os repositórios digitais devem contar com uma "Política de Indexação", que deve estar integrada aos outros aspectos do sistema de recuperação da informação. Segundo esses autores, uma Política de Indexação bem formulada pode garantir maior clareza, consistência e eficiência no processo de indexação dos documentos.

Políticas, de forma geral, têm o objetivo de padronizar comportamentos e operações, garantindo consistência e eficiência, além de orientar os agentes envolvidos na gestão dos sistemas de informação. Elas também têm um papel fundamental na gestão de riscos, ao minimizar erros e comportamentos inadequados, e na promoção da cultura organizacional, refletindo os valores e os princípios da instituição responsável pelo sistema.

Assim como em bibliotecas físicas convencionais, com suas usuais Políticas de desenvolvimento e gestão de acervo, de catalogação e de visitação e pesquisa, Políticas de Indexação podem colaborar para um posicionamento transparente do sistema em relação aos procedimentos adotados na indexação, sendo a sua formalização em documentos uma ação associada à sua continuidade e ao seu aperfeiçoamento; possuem, ainda, impacto político, visto que explicitam decisões da entidade que a originou.

Essas Políticas podem ser a base para que outros instrumentos norteadores da representação temática da informação sejam elaborados, a exemplo dos, já mencionados, Manuais de Indexação, Guias de Indexação, Políticas de Autoarquivamento, Manuais de Autoarquivamento e Guias de Autoarquivamento e Tutoriais para Autoarquivamento. Neste ponto, alerta-se para que não sejam confundidos os conceitos de "Política de Indexação" e dos documentos supracitados, visto que estes últimos são documentos derivados que podem conter, mais objetivamente, diretrizes de uma Política de Indexação; correspondendo a um dos meios pelo qual a Política de Indexação de um sistema de informação poderá ser observada.

Historicamente, as Políticas de Indexação foram inicialmente direcionadas aos profissionais responsáveis pela criação de registros bibliográficos em bibliotecas físicas e, mais tarde, digitais. No cenário digital e colaborativo atual, essas Políticas têm se tornado cada vez mais relevantes também para a comunidade produtora de informação, especialmente em contextos que envolvem o autoarquivamento de dados de pesquisa.

Para Oliveira *et al.* (2021, p. 149, grifo nosso), “[...] a política de indexação é tida justamente como um conjunto de orientações reguladoras da indexação e precisa ser considerada *em todos* os sistemas de informação nos quais o ato de indexar está presente”. Portanto, essas políticas não devem significar “[...] apenas uma formalidade para os profissionais, mas um guia útil que demonstra a direção” (Andrade; Sabbag, 2017, p. 82), prevendo também a atuação de autores científicos como indexadores especialistas.

A primeira discussão sobre a Política de Indexação na Ciência da Informação brasileira remonta a 1985, com o artigo de Marília Vidigal Carneiro, intitulado “Diretrizes para uma Política de Indexação”. Sua obra é considerada um marco no campo, propondo princípios fundamentais para a elaboração e manutenção dessas políticas.

De acordo com Carneiro (1985), de maneira ampla, Políticas de Indexação são ferramentas que têm como objetivo final garantir que determinada informação seja fornecida ao usuário no momento preciso e da maneira mais eficiente e econômica possível, através de um processo de indexação guiado. Neste sentido, seu intuito é a definição das variáveis que afetam o desempenho do sistema de informação, o estabelecimento de critérios e princípios que tornarão esse sistema mais eficiente, a racionalização dos processos e a consistência das operações envolvidas. Em sua elaboração, equipes de profissionais devem se atentar a alguns condicionantes do planejamento de quaisquer sistemas de recuperação da informação, sendo eles: a) as características e objetivos da organização; b) o perfil do usuário a ser atendido; e, c) a conjuntura dos recursos humanos, materiais e financeiros disponíveis para o fim (Carneiro, 1985). Fujita (2021) acrescenta que, para a obtenção de uma visão abrangente, antes de formular uma Política de Indexação é necessária a análise do

funcionamento do respectivo sistema e da recuperação da informação; após essa etapa, devem ser estabelecidas as diretrizes, os elementos e as variáveis que constituirão a Política.

Com os subsídios da literatura científica clássica dedicada às Políticas de Indexação, Fujita e Santos (2016, p. 5) determinaram que elas devem reunir

[...] elementos, variáveis, procedimentos, instrumentos e métodos de avaliação definidos pela instituição a que se destina em função de requisitos de organização, clientela, recursos financeiros, materiais e humanos. Essa política terá como *elementos*: cobertura de assuntos ou domínios de especialidade, documentos ou recursos de informação, estratégia de busca, o processo de indexação e a avaliação. O processo de indexação como elemento essencial da política é, por sua vez, influenciado pelas *variáveis* da recuperação da informação: escolha da linguagem, exaustividade, especificidade, revocação e precisão.

Por fim, as Políticas de Indexação devem ser flexíveis e adaptáveis às mudanças no ambiente em que são implementadas. A avaliação contínua dessas políticas é essencial para garantir sua adequação e efetividade, e deve ser um compromisso dos gestores e bibliotecários envolvidos no processo.

A representação temática em sistemas híbridos de recuperação informação, ou seja, aqueles em que o profissional da informação não é o único atuante na indexação de documentos e na sua inclusão no sistema, demanda diretrizes que atendam fundamentalmente às novas complexidades experienciadas, como as relativas à indexação de dados de pesquisa autoarquivados.

Com esta perspectiva, segue-se para o próximo passo desta pesquisa, que objetiva a análise de critérios norteadores contidos em diversos documentos que possam orientar a atribuição de palavras-chave no depósito de dados de pesquisa por autores científicos em repositórios digitais do Brasil, Espanha e Portugal. Os resultados obtidos podem servir como subsídios para a formalização de procedimentos mais estruturados e transparentes no contexto do autoarquivamento de dados de pesquisa, colaborando com a recuperação da informação.

### 3 PROCEDIMENTOS METODOLÓGICOS

Esta seção descreve os procedimentos metodológicos adotados para a realização desta pesquisa, fundamentando-se nas contribuições teóricas de Lakatos e Marconi (2007) e Gil (2011).

#### 3.1 CARACTERIZAÇÃO DA PESQUISA

A pesquisa em questão é de natureza aplicada, com abordagem qualitativa. Seus objetivos são de caráter descritivo e exploratório, e os procedimentos de coleta de dados envolvem fontes bibliográficas e documentais.

A *natureza aplicada* da pesquisa justifica-se pela busca de soluções para problemas práticos no campo da Ciência da Informação. Essa abordagem visa contribuir para a resolução de questões específicas dentro dessa área científica (Gerhardt; Silveira, 2009).

A *abordagem qualitativa* envolve a descrição de fenômenos, caracterizando sua ocorrência e relacionando-os com outros fatores (Casarin; Casarin, 2012, p. 33), o que corresponde aos objetivos da presente pesquisa. O foco são os aspectos qualitativos das diretrizes para a indexação em repositórios de dados de pesquisa do Brasil, Espanha e Portugal, tendo como intuito a análise dos critérios apresentados para que autores científicos realizem autoarquivamento desses recursos. Assim, os dados quantitativos e a respectiva análise quantitativa cumprem o papel de subsidiar a análise qualitativa.

A pesquisa é *exploratória*, pois busca proporcionar uma compreensão mais profunda sobre o problema de pesquisa, utilizando o conhecimento existente no campo da Ciência da Informação (Gil, 2011). A exploração bibliográfica realizada durante o ano de 2023 e o primeiro semestre de 2024 foi fundamental para a elaboração da seção introdutória e do referencial teórico. Essa revisão de literatura

possibilitou a construção de análises e reflexões sobre o tema, além de que forneceu a base para a discussão dos dados obtidos.

Os materiais bibliográficos consultados contemplaram fontes de informação dedicadas especialmente à questão da *Ciência Aberta*, aos *dados de pesquisa abertos*, aos aspectos gerais dos *repositórios digitais* e aos *repositórios de dados de pesquisa*, à prática do *autoarquivamento* nestes sistemas, ao emprego de *metadados temáticos* na indexação em sistemas de recuperação da informação, e ao papel de *diretrizes para a indexação* em sistemas que adotam o autoarquivamento. Recorreu-se ao acesso a publicações científicas disponíveis em *bases de dados nacionais* como a Base de Dados Referencial de Artigos de Periódicos em Ciência da Informação<sup>23</sup> (BRAPCI) e a Biblioteca Digital Brasileira de Teses e Dissertações<sup>24</sup> (BDTD); nos Anais do Encontro Nacional de Pesquisa em Ciência da Informação (ENANCIB) e em *bases de dados internacionais* disponíveis para consulta a partir do Portal de Periódicos CAPES<sup>25</sup>, tais como *Library & Information Science Abstracts (LISA)*, *Scientific Electronic Library Online (SciELO)*, *Scopus* e *Web of Science*.

Por fim, a pesquisa se caracteriza como *descritiva* ao recorrer à diretrizes para a indexação formalizadas em repositórios de dados de pesquisa, analisando-as mediante o emprego da técnica de Análise de Conteúdo, de Laurence Bardin (1977) - composta pela *pré-análise*; pela *exploração do material*; e pelo *tratamento dos resultados e interpretação*.

### 3.2 UNIVERSO DA PESQUISA

O universo da pesquisa é composto por repositórios de dados de pesquisa que adotam a prática de autoarquivamento e oferecem documentos com diretrizes para a atribuição de palavras-chave pelos autores científicos no depósito de seus dados, com acesso integral aos materiais.

---

<sup>23</sup> Disponível em: <https://www.brapci.inf.br/>.

<sup>24</sup> Disponível em: <https://bdtd.ibict.br/vufind/>.

<sup>25</sup> Disponível em: <https://www-periodicos-capes-gov-br.ez1.periodicos.capes.gov.br/index.php?>.

Para a seleção sistemática desses repositórios digitais, foi utilizado o diretório global de repositórios de dados de pesquisa chamado re3data, uma ferramenta eletrônica que direciona os usuários para essas infraestruturas e fornece informações detalhadas sobre elas.

Lançado em 2012 com financiamento da *German Research Foundation (DFG)*, o re3data visa promover uma cultura de compartilhamento de dados, aumentar o acesso e a visibilidade de conjuntos de dados de pesquisa (Pavão *et al.*, 2018), oferecendo uma visão geral de repositórios internacionais. Contudo, vale destacar que, embora seja uma ferramenta relevante para a Ciência Aberta, o re3data não garante, necessariamente, ampla acessibilidade aos dados, já que repositórios registrados podem restringir o acesso aos dados de pesquisa. De acordo com a seção de perguntas frequentes do re3data, os termos de uso e acesso definidos pela equipe gestora “[...] são provavelmente a questão mais importante de um repositório de dados de pesquisa”, especialmente no que diz respeito aos repositórios indexados no diretório.

No re3data, um repositório de dados de pesquisa é incluído apenas quando atende aos requisitos mínimos de uma Política específica. Esta Política aborda aspectos como o modo de acesso aos dados e aos repositórios, além de exigir que os termos de uso sejam claramente especificados nos *sites* dos sistemas. Um repositório deve ter foco exclusivo em dados de pesquisa para ser elegível para registro no diretório.

Para definir a acessibilidade global de um repositório de dados de pesquisa, termos de uso/acesso podem ser divididos em três categorias que definem diferentes níveis de acesso: 1) *acesso ao repositório*: refere-se ao acesso ao *site* do repositório; 2) *acesso aos dados*: refere-se ao acesso aos conjuntos de dados depositados no repositório; e 3) *upload de dados*: refere-se à possibilidade de enviar conjuntos de dados ao repositório, tanto por usuários internos quanto externos à(s) instituição(ões) responsável(is).

O acesso a cada nível pode ser aberto, restrito e/ou fechado. *Aberto*, quando não ocorre quaisquer barreiras de acesso a quaisquer usuários aos repositórios e

aos conjuntos de dados depositados; *restrito*, quando o repositório é disponível de maneira plenamente livre apenas a determinada comunidade, entretanto, com a possibilidade de que outros usuários superem barreiras de acesso com permissões específicas; e *fechado*, quando usuários externos à determinada comunidade principal, de maneira alguma, podem superar barreiras de acesso (re3data, s/d).

O re3data também auxilia pesquisadores e organizações na escolha de repositórios para o depósito de dados de pesquisa<sup>26</sup>. Nesse contexto, ele é recomendado nas "Diretrizes sobre Acesso Aberto a Publicações Científicas e Dados de Pesquisa no Horizonte 2020", da Comissão Europeia. Além disso, editoras e periódicos, como *Copernicus Publications*<sup>27</sup>, *PeerJ*<sup>28</sup>, *Springer*<sup>29</sup> e *Nature's Scientific Data*<sup>30</sup>, referem-se ao re3data em suas políticas editoriais como ferramenta para a fácil identificação de repositórios apropriados para armazenar dados de pesquisa.

Neste estudo, a escolha do re3data como diretório para acesso aos repositórios digitais foi motivada principalmente pela sua facilidade de uso, em comparação com outras ferramentas como o *FAIRsharing*<sup>31</sup>, devido à sua interface mais intuitiva e ao conteúdo mais acessível. Além disso, optou-se por usar exclusivamente o re3data, pois a combinação com outros diretórios extrapolaria o número viável de componentes para o universo da pesquisa, dificultando a análise.

Até janeiro de 2024, o re3data registrava cerca de 3.200 repositórios com foco em dados de pesquisa, incluindo aqueles que também armazenam monografias, como Teses e Dissertações. As áreas temáticas dos repositórios cadastrados são amplamente diversificadas, refletindo a diversidade de dados gerados no contexto da pesquisa científica global.

---

<sup>26</sup>Conforme Pavão *et al.* (2018, p. 1), para se registrar no re3data um repositório de dados de pesquisa deve obedecer a certos requisitos, a saber: ser administrado por uma entidade legal como, por exemplo, uma biblioteca, uma universidade etc.; esclarecer as condições de acesso aos dados e ao repositório, bem como aos termos de uso; e ter foco em dados de pesquisa.

<sup>27</sup> Disponível em: <https://publications.copernicus.org/>.

<sup>28</sup> Disponível em: <https://peerj.com/>.

<sup>29</sup> Disponível em: <https://www.springer.com/br>

<sup>30</sup> Disponível em: <https://www.nature.com/sdata/>.

<sup>31</sup> Disponível em: <https://fairsharing.org/>.

Na realização deste estudo, o acesso aos repositórios se deu em janeiro de 2024. Inicialmente, com a utilização do índice de país do referido diretório, foram filtrados todos os registros de repositórios digitais do Brasil, Espanha e Portugal. Nesse sentido, conforme documentado no Quadro do Apêndice A, foram quantificados um total de 89 (oitenta e nove) registros de repositórios digitais: 21 (vinte e um) vinculados ao Brasil; 55 (cinquenta e cinco) vinculados à Espanha; 15 (quinze) vinculados à Portugal; e 2 (dois) vinculados concomitantemente à Espanha e a Portugal (o *European Archive of Historical Earthquake Data - AHEAD* e o *Antimicrobial Combination Networks*). A ordenação dos registros desses sistemas no Quadro do Apêndice A corresponde à ordenação dos mesmos no momento em que foram recuperados.

A tentativa de acesso às interfaces dos 89 (oitenta e nove) repositórios digitais permitiu constatar a indisponibilidade de 3 (três), não estando eles em funcionamento no momento da coleta de dados para esta pesquisa; são eles: o “Banco de dados de exploração e produção”, vinculado ao Brasil; o “*Social Impact of Science*” (SIOR), vinculado à Espanha; e o “*Kinetic models of biological systems*” (KiMoSys), vinculado à Portugal. De antemão, esses repositórios digitais foram desconsiderados para a análise.

Em atenção ao foco desta pesquisa, ou seja, repositórios de dados de pesquisa que, *restritivamente* armazenam este tipo de recurso informacional, foram também desconsiderados quaisquer *repositórios digitais mistos*, ou seja, aqueles que armazenavam outros tipos de trabalhos científicos, como Teses e Dissertações. Dessa forma, foram excluídos do universo da pesquisa outros 6 (seis) repositórios digitais, sendo eles: o “*Arias Montano*”, o “*CEU Institutional Repository*”, o “*citaREA*”, o “*Repositori Institucional Universitat Rovira i Virgili*”, o “RIUMA” e o “UCREA”; relacionados à Espanha.

Em um segundo momento, mediante análise dos ambientes informacionais dos 80 (oitenta) repositórios digitais restantes, foram selecionados os que funcionam como sistemas de autoarquivamento e que, adicionalmente, ofereciam documentos com diretrizes para a atribuição de palavras-chave aos dados, sendo possível seu acesso no momento da análise (conforme o Apêndice A). Nesta etapa, constatou-se que, na maioria dos casos, os documentos apresentando diretrizes

para a indexação de dados de pesquisa não estavam facilmente localizáveis nos ambientes informacionais dos repositórios digitais, sendo necessárias buscas exaustivas para identificá-los; fato que, na prática do autoarquivamento, pode desencorajar o autor científico e dificultar a sua adesão para a representação temática dos seus dados.

Por fim, através do processo de seleção anteriormente detalhado, foi composto um universo de pesquisa de 21 (vinte e um) repositórios digitais habilitados para análise de suas diretrizes; sendo 12 (doze) brasileiros, 3 (três) espanhóis e 6 (seis) portugueses, conforme apresentados no Quadro 7. Como se verifica, boa parte desses repositórios digitais se relacionam com, pelo menos, um dos países envolvidos neste estudo, Brasil, Espanha e Portugal, e ao mesmo tempo com outros, em sentido de parceria institucional internacional.


No Quadro 7, além dos nomes dos repositórios digitais e dos países que administram esses sistemas, estão dispostas as tipologias<sup>32</sup> dos mesmos, além da Área do Conhecimento a que se relacionam<sup>33</sup>.

---

<sup>32</sup> Institucional, Disciplinar, Multidisciplinar e/ou Orientada por projetos (Pampel *et al.*, 2013).



<sup>33</sup> De acordo com a Tabela de Áreas do Conhecimento definidas pelo CNPQ. Disponível em: <https://lattes.cnpq.br/web/dgp/arvore-do-conhecimento>. Acesso em: 26 out. 2024.

**Quadro 7** - Repositórios de dados de pesquisa do Brasil, Espanha e Portugal que adotam o autoarquivamento (re3data).

	Nome do repositório	Descrição	País (es) a que se vincula	Tipologia (s)	Área (s) do Conhecimento/
   BRASIL (BR)   QUANTIDADE DE REGISTROS: 12					
01	<b>Aleia</b>	Repositório oficial do IBICT que custodia, publica, dissemina, preserva e compartilha conjuntos de dados da comunidade científica do IBICT (constituída por seus servidores, pesquisadores visitantes, bolsistas e alunos do programa de pós-graduação). Tem como objetivo promover, apoiar e facilitar a adoção do acesso aberto aos dados de pesquisa, fomentar novas pesquisas e promover uma Ciência Aberta. A criação do Aleia foi motivada pelo Compromisso 3, do 4º Plano de Ação Nacional que visa estabelecer mecanismos de governança de dados científicos para o avanço da Ciência Aberta no Brasil. Atualmente, o Aleia tem como finalidade tornar acessível e dar visibilidade aos dados de pesquisa elaborados pela comunidade científica Ibict; garantir a reprodutibilidade e reuso dos dados de pesquisa da comunidade científica institucional do IBICT; funcionar como ferramenta para promover a transparência e por consequência a confiança e a solidez das pesquisas desenvolvidas, maximizando o retorno de aplicação de recurso público; viabilizar a preservação digital de dados de pesquisa, colaborando para evitar a perda e má gestão dos dados de pesquisa, orientando e incentivando os autores na organização e planejamento do ciclo de vida do dado de pesquisa.	<i>Brasil (Instituto Brasileiro de Informação em Ciência e Tecnologia)</i>	Institucional; disciplinar	- CIÊNCIAS HUMANAS - CIÊNCIAS SOCIAIS APLICADAS
02	<b>Arca Dados</b>	Repositório oficial de dados para pesquisa da Fiocruz. Tem a missão de ser a plataforma para arquivar, publicar, disseminar, preservar e compartilhar os dados digitais para pesquisa produzidos pela comunidade Fiocruz e/ou em parceria com outros Institutos ou órgãos de pesquisa, seguindo os princípios FAIR.	<i>Brasil (Fundação Oswaldo Cruz); outros</i>	Institucional; multidisciplinar	- CIÊNCIAS BIOLÓGICAS - CIÊNCIAS EXATAS E DA TERRA - ENGENHARIAS - CIÊNCIAS HUMANAS - CIÊNCIAS SOCIAIS APLICADAS
03	<b>Dados Abertos de Pesquisas</b>	Iniciativa da Diretoria de Pós-Graduação, Pesquisa e Inovação do Instituto Federal Goiano – Campus Urutaí. Segue a filosofia de Ciência Aberta, para expansão e valorização da pesquisa científica, visando disponibilizar dados oriundos de observações e experimentação técnico-científicas, garantindo que os seus autores, pesquisadores e estudantes recebam todo o merecido crédito enquanto agentes geradores de dados.	<i>Brasil (Instituto Federal de Educação Ciência e Tecnologia Goiano); outros</i>	Institucional; multidisciplinar	- CIÊNCIAS AGRÁRIAS - CIÊNCIAS BIOLÓGICAS - CIÊNCIAS EXATAS E DA TERRA - ENGENHARIAS - ENGENHARIAS
04	<b>Deposita dados</b>	O Repositório Comum de Dados de Pesquisa (Deposita Dados) integra e dissemina, conjuntos de dados de pesquisa de pesquisadores brasileiros vinculados a instituições científicas que ainda não possuem seus repositórios de dados de pesquisa e/ou de pesquisadores brasileiros que executaram seus conjuntos de dados por meio de colaboração científica em instituições estrangeiras de ensino e pesquisa. O acesso a esses conjuntos de dados é livre de quaisquer custos. O Deposita Dados funciona como ferramenta para promover a transparência e	<i>Brasil (Instituto Brasileiro de Informação em Ciência e Tecnologia)</i>	Institucional; multidisciplinar	- CIÊNCIAS BIOLÓGICAS - CIÊNCIAS EXATAS E DA TERRA - ENGENHARIAS - CIÊNCIAS HUMANAS

	Nome do repositório	Descrição	País (es) a que se vincula	Tipologia (s)	Área (s) do Conhecimento/
		possibilitar o reuso dos dados, aumentando assim a reprodutibilidade e por consequência a confiança e a solidez das pesquisas desenvolvidas, maximizando o retorno de aplicação de recursos públicos.			- CIÊNCIAS SOCIAIS APLICADAS
05	<b>GLOBE</b>	Ambiente colaborativo <i>on-line</i> que permite aos pesquisadores de mudanças territoriais compartilhar, comparar e integrar estudos locais e regionais com dados globais para avaliar a relevância global de seu trabalho. Projeto financiado pela <i>National Science Foundation</i> conduzido em colaboração com o <i>Global Land Project</i> , uma equipe de pesquisadores da Universidade de Maryland e estudantes de Geografia e Sistemas Ambientais, Ciência da Computação e Engenharia Elétrica e Sistemas de Informação.	Estados Unidos; <i>Brasil (Instituto Nacional de Pesquisa Espacial, Projeto Global de Terras)</i>	Disciplinar	- CIÊNCIAS BIOLÓGICAS - CIÊNCIAS EXATAS E DA TERRA
06	<b>Maenduar</b>	Maenduar em tupi significa “lembrar”. O repositório Maenduar, criado pelo Laboratório em Rede de Humanidades Digitais (LARHUD) é um repositório institucional (IBICT) dedicado a dados de pesquisa em Humanidades Digitais e Humanidades. O repositório engloba a produção dos membros e parceiros do LARHUD e estende-se ao espaço público como repositório de depósitos públicos. Pesquisadores, professores, estudantes vinculados a quaisquer instituições poderão depositar seus dados de pesquisa, entendendo que o Maenduar se torna, a partir desta ação, um repositório público, nacional, dedicado à produção de conhecimento nas Ciências Humanas diante do desenvolvimento do digital cultura em todas as áreas do mundo. mundo social.	<i>Brasil (Instituto Brasileiro de Informação em Ciência e Tecnologia; Instituto Brasileiro de Informação em Ciência e Tecnologia, Laboratório em Rede de Humanidades Digitais)</i>	Institucional; disciplinar	- CIÊNCIAS HUMANAS - CIÊNCIAS SOCIAIS APLICADAS
07	<b>PPBio Data Repository</b>	Programa de Pesquisa em Biodiversidade - criado para aprofundamento de estudos sobre biodiversidade no Brasil. Faz uso de <i>software</i> de gerenciamento de dados 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 <i>Extensible Markup Language (XML)</i> .	<i>Brasil (INCT Centro de Estudos Integrados da Biodiversidade Amazônica); outros</i>	Disciplinar	- CIÊNCIAS BIOLÓGICAS
08	<b>Rede IBICT Cariniana Dataverse</b>	Dedicado ao arquivamento de dados de pesquisa por colaboradores das instituições participantes da rede Cariniana.	<i>Brasil (Instituto Brasileiro de Informação em Ciência e Tecnologia, Rede Cariniana; Ministério da Ciência,</i>	Institucional; multidisciplinar	- CIÊNCIAS BIOLÓGICAS - CIÊNCIAS EXATAS E DA TERRA - ENGENHARIAS - CIÊNCIAS HUMANAS - CIÊNCIAS SOCIAIS APLICADAS

	Nome do repositório	Descrição	País (es) a que se vincula	Tipologia (s)	Área (s) do Conhecimento/
			<i>Tecnologia e Inovações</i>		
09	<b>Redape</b>	Gerido pela Secretaria de Pesquisa e Desenvolvimento (SPD) em parceria com a Embrapa Agricultura Digital. Foi desenvolvido de maneira colaborativa por representantes de várias unidades. A proposta é que o Redape seja uma plataforma dinâmica, que contemple a diversidade e complexidade das ciências de dados e atenda de forma simples e precisa as necessidades das equipes de pesquisa da Embrapa.	<i>Brasil (Empresa Brasileira de Pesquisa Agropecuária)</i>	Institucional; disciplinar	<ul style="list-style-type: none"> <li>- CIÊNCIAS AGRÁRIAS</li> <li>- CIÊNCIAS BIOLÓGICAS</li> </ul>
10	<b>REDU</b>	Repositório oficial da Universidade de Campinas, para o depósito, preservação, compartilhamento, reutilização e reprodutibilidade de todos os dados de pesquisa gerados por sua comunidade.	<i>Brasil (Universidade de Campinas)</i>	Institucional; multidisciplinar	<ul style="list-style-type: none"> <li>- CIÊNCIAS AGRÁRIAS</li> <li>- CIÊNCIAS BIOLÓGICAS</li> <li>- CIÊNCIAS DA SAÚDE</li> <li>- CIÊNCIAS EXATAS E DA TERRA</li> <li>- ENGENHARIAS</li> <li>- CIÊNCIAS HUMANAS</li> <li>- CIÊNCIAS SOCIAIS APLICADAS</li> <li>- LINGÜÍSTICA, LETRAS E ARTES</li> </ul>
11	<b>Repositório de Dados de Pesquisa Unifesp</b>	Plataforma de armazenamento, preservação e acesso a dados de pesquisas voltados para a comunidade acadêmica da Universidade Federal de São Paulo.	<i>Brasil (Universidade Federal de São Paulo)</i>	Institucional; multidisciplinar	<ul style="list-style-type: none"> <li>- CIÊNCIAS AGRÁRIAS</li> <li>- CIÊNCIAS BIOLÓGICAS</li> <li>- CIÊNCIAS DA SAÚDE</li> <li>- CIÊNCIAS EXATAS E DA TERRA</li> <li>- ENGENHARIAS</li> <li>- CIÊNCIAS HUMANAS</li> <li>- CIÊNCIAS SOCIAIS APLICADAS</li> <li>- LINGÜÍSTICA, LETRAS E ARTES</li> </ul>
12	<b>SciELO Data</b>	Repositório para depósito, preservação e divulgação de dados de pesquisas provenientes de artigos submetidos e aprovados para publicação, já publicados em periódicos da Rede SciELO ou depositados no SciELO Preprints. O SciELO Data oferece à comunidade de autores e editores da revista SciELO e aos usuários do SciELO Preprint um repositório de dados de pesquisa alinhados às melhores práticas e padrões internacionais para gerenciamento de dados de pesquisa. Os periódicos SciELO operam um repositório para indexação, preservação e divulgação de dados relacionados aos artigos que publicam.	<i>Brasil (Biblioteca Eletrônica Científica Online)</i>	Institucional; multidisciplinar	<ul style="list-style-type: none"> <li>- CIÊNCIAS AGRÁRIAS</li> <li>- CIÊNCIAS BIOLÓGICAS</li> <li>- CIÊNCIAS DA SAÚDE</li> <li>- CIÊNCIAS EXATAS E DA TERRA</li> <li>- ENGENHARIAS</li> <li>- CIÊNCIAS HUMANAS</li> </ul>

	Nome do repositório	Descrição	País (es) a que se vincula	Tipologia (s)	Área (s) do Conhecimento/
					<ul style="list-style-type: none"> <li>- CIÊNCIAS SOCIAIS APLICADAS</li> <li>- LINGUÍSTICA, LETRAS E ARTES</li> </ul>
   ESPANHA (ES)   QUANTIDADE DE REGISTROS: 03					
01	<b>CORA. Repositori de Dades de Recerca</b>	Repositório de dados abertos, com curadoria e FAIR que abrange todas as disciplinas académicas. CORA. Repositori de dades de Recerca é um serviço partilhado prestado pelas instituições catalãs participantes (Universidades e Centros de Investigação CERCA). O repositório é gerenciado pelo CSUC e a infraestrutura técnica é baseada no aplicativo Dataverse, desenvolvido por desenvolvedores e usuários internacionais liderados pela Universidade de Harvard.	<i>Espanha (vários)</i>	Institucional; multidisciplinar	<ul style="list-style-type: none"> <li>- CIÊNCIAS AGRÁRIAS</li> <li>- CIÊNCIAS BIOLÓGICAS</li> <li>- CIÊNCIAS DA SAÚDE</li> <li>- CIÊNCIAS EXATAS E DA TERRA</li> <li>- ENGENHARIAS</li> <li>- CIÊNCIAS HUMANAS</li> <li>- CIÊNCIAS SOCIAIS APLICADAS</li> <li>- LINGUÍSTICA, LETRAS E ARTES</li> </ul>
02	<b>ioChem-BD</b>	Repositório digital para gerenciar ativos digitais de Química Computacional e Ciência de Materiais. A plataforma segue a filosofia das bases de dados cristalográficas, os princípios e necessidades do acesso aberto, e resolve problemas de reprodutibilidade e rastreabilidade.	<i>Espanha (Barcelona Supercomputing Center; Institute of Chemical Research of Catalonia; Universitat Rovira i Virgili)</i>	Institucional; disciplinar	<ul style="list-style-type: none"> <li>- CIÊNCIAS EXATAS E DA TERRA</li> </ul>
03	<b>RiuNet</b>	Destina-se a guardar a produção da comunidade universitária, pessoal ou institucional, em acervos. Estes podem ser constituídos por diversos tipos de documentos como Objetos de aprendizagem (Polimedia, laboratórios virtuais e artigos educativos), Teses, artigos de periódicos, mapas, trabalhos académicos, trabalhos criativos, património institucional, multimídia, material didático, produção institucional, eletrônicos periódicos, anais de conferências e dados de pesquisas.	<i>Espanha (Universitat Politècnica de València; Servicio de Biblioteca y Documentación Científica)</i>	Institucional; multidisciplinar	<ul style="list-style-type: none"> <li>- CIÊNCIAS BIOLÓGICAS</li> <li>- CIÊNCIAS EXATAS E DA TERRA</li> <li>- ENGENHARIAS</li> </ul>
   PORTUGAL (PT)   QUANTIDADE DE REGISTROS: 06					
01	<b>APIS</b>	Infraestrutura científica para as ciências sociais cuja principal atividade consiste na preservação e disseminação de dados de investigação. Com sede no Instituto de Ciências Sociais da	<i>Portugal (Fundação para a</i>	Disciplinar	<ul style="list-style-type: none"> <li>- CIÊNCIAS HUMANAS</li> </ul>

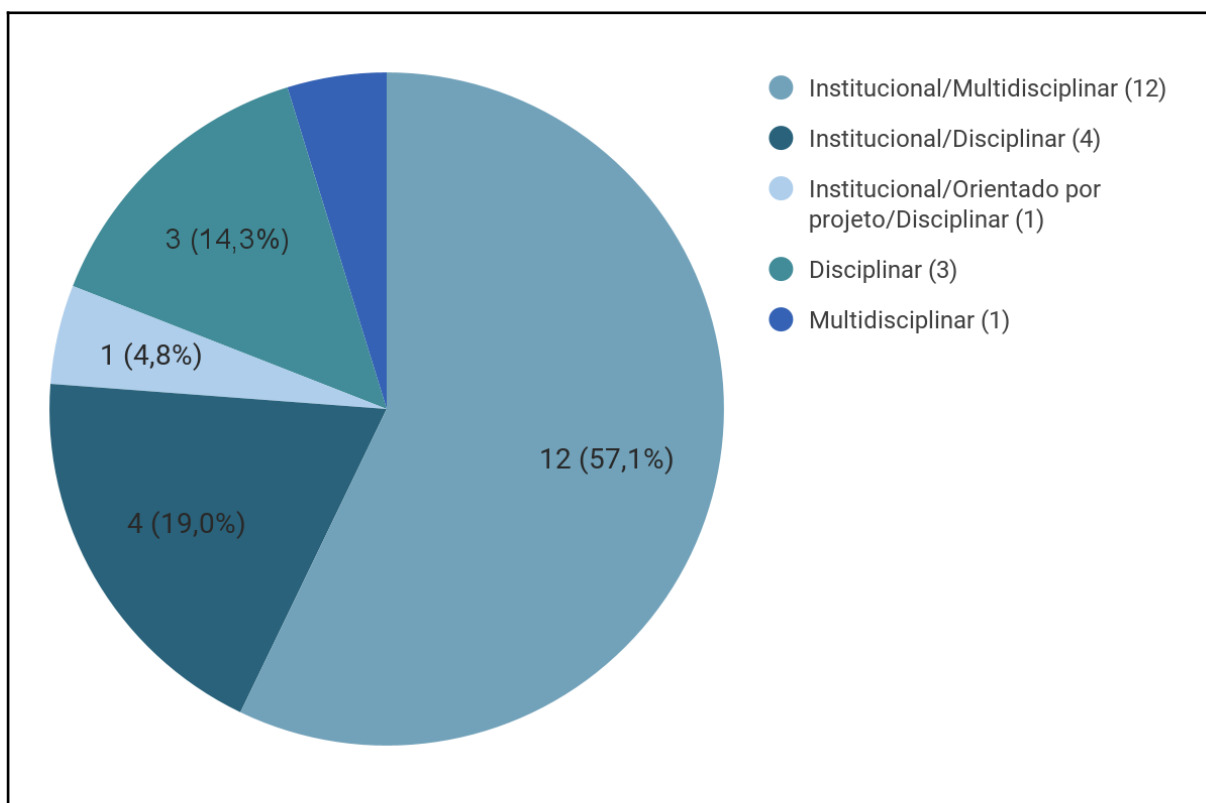
	Nome do repositório	Descrição	País (es) a que se vincula	Tipologia (s)	Área (s) do Conhecimento/
		Universidade de Lisboa, o arquivo promove a angariação e a partilha de dados de investigação em formato digital com vista à sua consulta pública, análise secundária e uso pedagógico. Do seu acervo fazem parte um conjunto de bases de dados resultantes de projetos de investigação realizados pela comunidade científica nacional.	<i>Ciência e a Tecnologia; Universidade de Lisboa, Instituto de Ciências Sociais)</i>		- CIÊNCIAS SOCIAIS APLICADAS
02	<b>DadosIPB</b>	Tem como objetivo armazenar, preservar e disponibilizar dados de pesquisas gerados pela comunidade académica, em acesso aberto – tão aberto quanto possível, tão fechado quanto necessário.	<i>Portugal (Instituto Politécnico de Bragança)</i>	Institucional; multidisciplinar	- CIÊNCIAS AGRÁRIAS - CIÊNCIAS BIOLÓGICAS - CIÊNCIAS DA SAÚDE - CIÊNCIAS EXATAS E DA TERRA - ENGENHARIAS - CIÊNCIAS HUMANAS - CIÊNCIAS SOCIAIS APLICADAS - LINGUÍSTICA, LETRAS E ARTES
03	<b>DataRepositoriUM</b>	Partilha, publica e gerencia dados das unidades de investigação da Universidade do Minho.	<i>Portugal (Universidade de Minho)</i>	Institucional; multidisciplinar	- CIÊNCIAS AGRÁRIAS - CIÊNCIAS BIOLÓGICAS - CIÊNCIAS DA SAÚDE - CIÊNCIAS EXATAS E DA TERRA - ENGENHARIAS - CIÊNCIAS HUMANAS - CIÊNCIAS SOCIAIS APLICADAS - LINGUÍSTICA, LETRAS E ARTES
04	<b>DUnAs</b>	Repositório institucional de dados de investigação da Universidade de Aveiro. Este repositório tem como objetivo partilhar, arquivar, preservar, citar, acessar e explorar dados de pesquisa produzidos nas atividades de pesquisa científica da universidade. DUnAs está alinhado com os Princípios Orientadores FAIR para gestão e administração de dados científicos.	<i>Portugal (Universidade de Aveiro)</i>	Institucional; multidisciplinar	- CIÊNCIAS AGRÁRIAS - CIÊNCIAS BIOLÓGICAS - CIÊNCIAS DA SAÚDE - CIÊNCIAS EXATAS E DA TERRA - ENGENHARIAS - CIÊNCIAS HUMANAS - CIÊNCIAS SOCIAIS APLICADAS

	Nome do repositório	Descrição	País (es) a que se vincula	Tipologia (s)	Área (s) do Conhecimento/
					- LINGUÍSTICA, LETRAS E ARTES
05	<b>Perdigao Field Experiment</b>	O projeto de campo da Perdigão faz parte do maior programa plurianual conjunto dos Estados Unidos da América e da Europa em Portugal. O projeto é parcialmente financiado pela ERANET+ da União Europeia para fornecer ao setor da energia eólica capacidades de mapeamento de recursos mais detalhadas na forma de um novo atlas eólico digital da UE. Um dos principais objetivos do projeto do campo Perdigão é quantificar erros de modelos de recursos eólicos em relação a um conjunto de dados de referência coletados em terrenos complexos. A participação dos Estados Unidos da América complementarará esta atividade, identificando as fraquezas físicas e numéricas dos modelos e desenvolvendo novos conhecimentos e métodos para superar essas deficiências	União Europeia; <i>Portugal</i> ( <i>Universidade do Porto, Faculdade de Engenharia</i> )	Orientado por projeto; Disciplinar	- CIÊNCIAS BIOLÓGICAS - CIÊNCIAS EXATAS E DA TERRA
06	<b>Repositório de Dados Científicos</b>	Plataforma de acesso gratuito a dados criados e utilizados no âmbito do trabalho de investigação de instituições nacionais. Caracteriza-se pela disponibilização de uma plataforma de repositório (DSpace) e suporte a toda a componente de manutenção de dados, como <i>backups</i> , monitorização, atualização, segurança, etc., mantendo assim os investigadores fora da preocupação destas tarefas. Por fim, o serviço SARDC pretende disponibilizar os dados depositados no repositório através do Portal RCAAP.	União Europeia; <i>Portugal</i> ( <i>Fundação para a Ciência e a Tecnologia;</i> <i>Project Partner;</i> <i>Repositórios Científicos de Acesso Aberto de Portugal;</i> <i>República Portuguesa</i> )	Multidisciplinar	- CIÊNCIAS AGRÁRIAS - CIÊNCIAS BIOLÓGICAS - CIÊNCIAS DA SAÚDE - CIÊNCIAS EXATAS E DA TERRA - ENGENHARIAS - CIÊNCIAS HUMANAS - CIÊNCIAS SOCIAIS APLICADAS - LINGUÍSTICA, LETRAS E ARTES

Fonte: elaboração própria (2024).

De modo complementar, a Figura 3 apresenta o panorama das tipologias dos repositórios investigados, baseando-se nas definições de Pampel *et al.* (2013).

**Figura 3** - Tipologias dos repositórios de dados de pesquisa.



Fonte: elaboração própria (2024).

Como se observa, 17 (dezesete) dos repositórios analisados (80,9%) são institucionais, sendo distribuídos da seguinte forma: 12 (doze [57,1%]) são repositórios multidisciplinares, 4 (quatro [19,0%]) são disciplinares e 1 (um [4,8%]) é orientado por projeto. Além disso, 3 (três) repositórios (14,3%) são exclusivamente disciplinares, sem restrições institucionais para o depósito de dados, e 1 (um) repositório (4,8%) é multidisciplinar, também sem limitações institucionais para depósito.

A seguir, será apresentado o procedimento adotado para a análise dos dados.

### 3.3 PROCEDIMENTO PARA ANÁLISE DOS DADOS

A análise dos dados foi realizada em *duas* etapas, conforme descrito a seguir.

A *primeira etapa* teve como objetivo *identificar e analisar a prevalência de critérios relacionados à atribuição de palavras-chave nos documentos localizados nos repositórios de dados de pesquisa incluídos no universo de estudo*. Alguns desses critérios foram definidos com base no referencial teórico da pesquisa, enquanto outros surgiram a partir da observação dos próprios documentos analisados. O Quadro 8 apresenta os 15 (quinze) critérios definidos e considerados para subsidiar a análise de forma sistemática.

**Quadro 8** - Critérios para a atribuição de palavras-chave.

Critério	Identificador
Apresenta definição própria para palavras-chave e/ou sua relevância para a representação de dados de pesquisa	critério 01
Recomenda o uso de um vocabulário controlado específico	critério 02
Recomenda o uso de vocabulário controlado a depender da escolha do depositante	critério 03
Indica a quantidade de termos/palavras-chave a serem incluídas	critério 04
Apresenta orientação para a ordenação de termos centrais e periféricos	critério 05
Apresenta orientação para o controle de termos polissêmicos; sinônimos; diferenciação de termos homógrafos e/ou hierarquização de termos	critério 06
Apresenta orientação para o emprego de letras maiúsculas e/ou sinais de pontuação e/ou acentuação gráfica	critério 07
Apresenta orientação para o emprego de termos singulares e plurais	critério 08
Apresenta orientação para a transcrição em língua estrangeira	critério 09
Apresenta orientação sobre a semantividade entre termos empregados	critério 10
Apresenta sugestão de termos no preenchimento de palavras-chave e/ou correção de erros ortográficos	critério 11
Disponibiliza Lista de perguntas frequentes e/ou Página de ajuda e/ou Tutorial	critério 12
Realiza a validação dos termos por pessoa bibliotecária	critério 13
Recomenda a repetição das palavras do título nas palavras-chave	critério 14
Recomenda a repetição das palavras do resumo nas palavras-chave	critério 15

Fonte: elaboração própria (2024).

A análise detalhada das diretrizes identificadas (conforme apresentado no quadro do Apêndice A) possibilitou a sistematização dos dados, culminando na

elaboração de um mapeamento dos critérios observados. A seguir, procedeu-se à mensuração da frequência de ocorrência desses critérios.


*Posteriormente*, foi visada a apresentação de considerações baseadas nos critérios para a elaboração de palavras-chave envolvidos neste estudo, que podem respaldar a adequação e o desenvolvimento de diretrizes (ou ainda de Políticas) que estabeleçam procedimentos mais estruturados e transparentes no contexto do autoarquivamento, contribuindo para a melhoria da recuperação da informação.



## 4 RESULTADOS E DISCUSSÃO

Esta seção apresenta e discute os principais resultados obtidos a partir da sistematização dos critérios identificados nos documentos disponibilizados pelos repositórios envolvidos nesta pesquisa. O processo de sistematização está refletido no Quadro 9, que indica a presença ou ausência de cada critério nos documentos analisados. As palavras "Sim" e "Não" foram utilizadas para sinalizar a presença e a ausência de cada critério, respectivamente, com destaque em cores verde e alaranjada. Os repositórios de dados de pesquisa estão organizados por país.

Vale ressaltar que o critério 03, “uso de vocabulário controlado a depender da escolha do depositante”, e o critério 14, “repetição das palavras do título nas palavras-chave”, de acordo com Terra *et al.* (2021), estão associados a práticas possivelmente negativas, conforme relatado em seu estudo. Explicações adicionais sobre esses pontos serão fornecidas adiante.

**Quadro 9 - Mapeamento dos critérios para a atribuição de palavras-chave na pesquisa documental.**

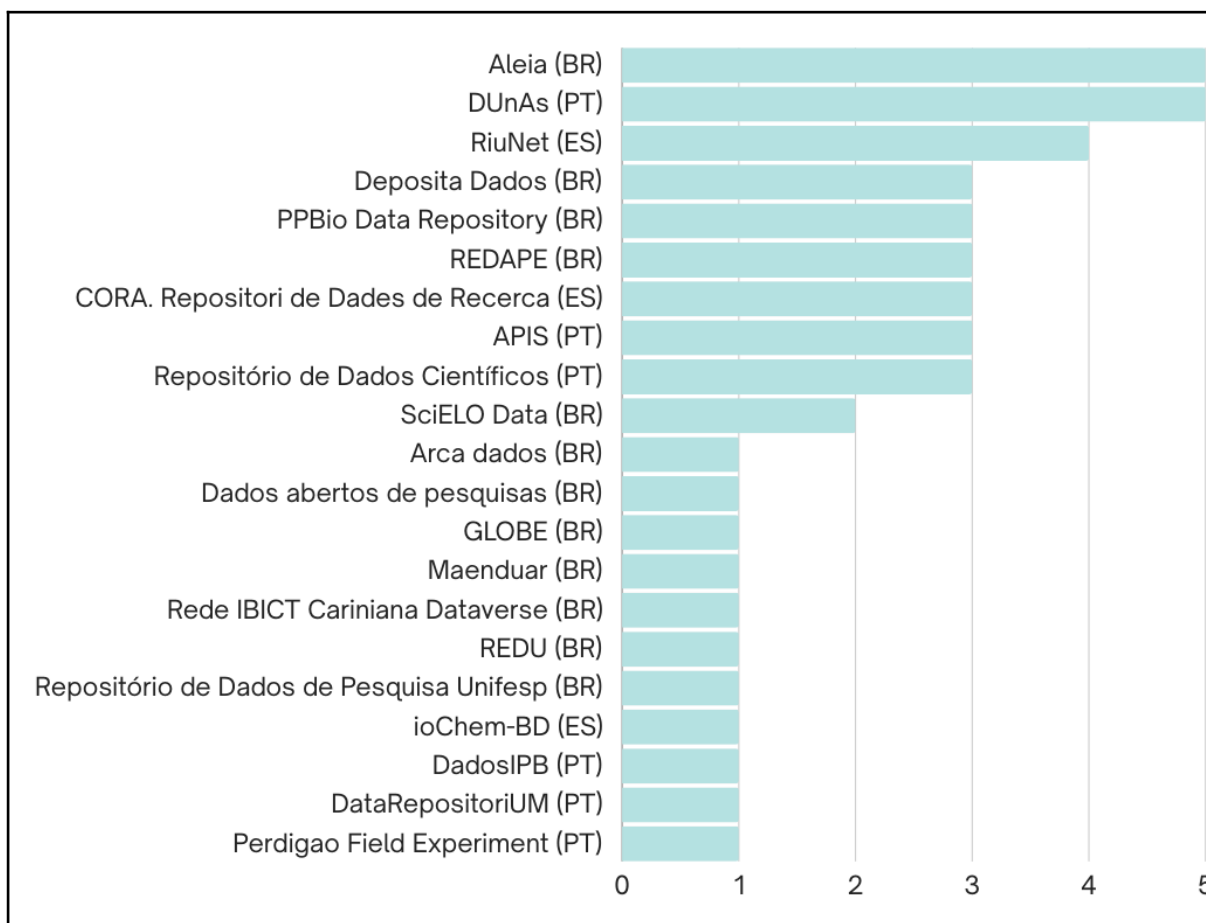
	Repositório	Tipo do documento com as diretrizes	Critérios															N. total de critérios por repositório
			critério 01	critério 02	critério 03	critério 04	critério 05	critério 06	critério 07	critério 08	critério 09	critério 10	critério 11	critério 12	critério 13	critério 14	critério 15	
   BRASIL (BR)   QUANTIDADE DE REPOSITÓRIOS: 12																		
01	Aleia	Cartilha	Sim	Não	Sim	Sim	Não	Não	Sim	Não	Não	Não	Informação inacessível	Não	Sim	Não	Não	05
02	Arca Dados	Guia	Não	Não	Sim	Não	Não	Não	Não	Não	Não	Não	Informação inacessível	Não	Informação indisponível	Não	Não	01
03	Dados Abertos De Pesquisas	Guia	Não	Não	Sim	Não	Não	Não	Não	Não	Não	Não	Informação inacessível	Não	Informação indisponível	Não	Não	01
04	Deposita Dados	Guias	Não	Não	Sim	Não	Não	Não	Sim	Não	Não	Não	Informação inacessível	Não	Sim	Não	Não	03
05	GLOBE	Tutoriais	Não	Não	Não	Não	Não	Não	Não	Não	Não	Sim	Informação inacessível	Não	Informação indisponível	Não	Não	01
06	Maenduar	Página para upload após login	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Sim	Não	Não	Não	Não	01
07	PPBio Data Repository	Cartilha; Manual	Sim	Sim	Não	Não	Não	Sim	Não	Não	Não	Não	Informação inacessível	Não	Informação indisponível	Não	Não	03
08	Rede IBICT Cariniana Dataverse	Guia	Não	Não	Sim	Não	Não	Não	Não	Não	Não	Não	Informação inacessível	Não	Informação indisponível	Não	Não	01
09	Redape	Guia; Orientações	Sim	Sim	Não	Não	Não	Não	Sim	Não	Não	Não	Informação inacessível	Não	Informação indisponível	Não	Não	03
10	REDU	Guia	Não	Não	Sim	Não	Não	Não	Não	Não	Não	Não	Informação inacessível	Não	Informação indisponível	Não	Não	01
11	Repositório de Dados de Pesquisa Unifesp	Guia	Não	Não	Sim	Não	Não	Não	Não	Não	Não	Não	Informação inacessível	Não	Informação indisponível	Não	Não	01

			Critérios															N. total de critérios por repositório
Repositório	Tipo do documento com as diretrizes	critério 01	critério 02	critério 03	critério 04	critério 05	critério 06	critério 07	critério 08	critério 09	critério 10	critério 11	critério 12	critério 13	critério 14	critério 15		
12	SciELO Data	Guia	Não	Não	Sim	Não	Não	Não	Não	Não	Não	Não	Informação inacessível	Não	Sim	Não	Não	02
   ESPANHA (ES)   QUANTIDADE DE REPOSITÓRIOS: 03																		
01	CORA. Repositori de Dades de Recerca	Guias	Sim	Não	Sim	Não	Não	Não	Sim	Não	Não	Não	Informação inacessível	Não	Informação indisponível	Não	Não	03
02	ioChem-BD	Guia	Não	Não	Sim	Não	Não	Não	Não	Não	Não	Não	Informação inacessível	Não	Informação indisponível	Não	Não	01
03	RiuNet	Página de ajuda	Não	Não	Sim	Sim	Sim	Não	Não	Não	Não	Não	Informação inacessível	Sim	Informação indisponível	Não	Não	04
   PORTUGAL (PT)   QUANTIDADE DE REPOSITÓRIOS: 06																		
01	APIS	Guia; Página de ajuda	Não	Sim	Não	Sim	Não	Não	Não	Não	Não	Não	Informação inacessível	Sim	Informação indisponível	Não	Não	03
02	DadosIPB	Guia	Não	Não	Sim	Não	Não	Não	Não	Não	Não	Não	Informação inacessível	Não	Informação indisponível	Não	Não	01
03	DataRepositoriUM	Guia	Não	Não	Sim	Não	Não	Não	Não	Não	Não	Não	Informação inacessível	Não	Informação indisponível	Não	Não	01
04	DUnAs	Guias	Sim	Não	Sim	Sim	Não	Não	Sim	Não	Não	Não	Informação inacessível	Não	Sim	Não	Não	05
05	Perdigao Field Experiment	Diretrizes; Instruções; Tesouro	Não	Sim	Não	Não	Não	Não	Não	Não	Não	Não	Informação inacessível	Não	Informação indisponível	Não	Não	01
06	Repositório de Dados Científicos	Guia; Página de ajuda	Não	Sim	Não	Sim	Não	Não	Não	Não	Não	Não	Informação inacessível	Sim	Informação indisponível	Não	Não	03

Fonte: elaboração própria (2024).

Considerando que esta pesquisa se direciona quantitativamente aos critérios identificados nos documentos de cada repositório investigado, a Figura 4 destaca a quantidade de critérios contemplados por cada repositório.

**Figura 4** - Quantidade de critérios contemplados por repositório de dados de pesquisa.



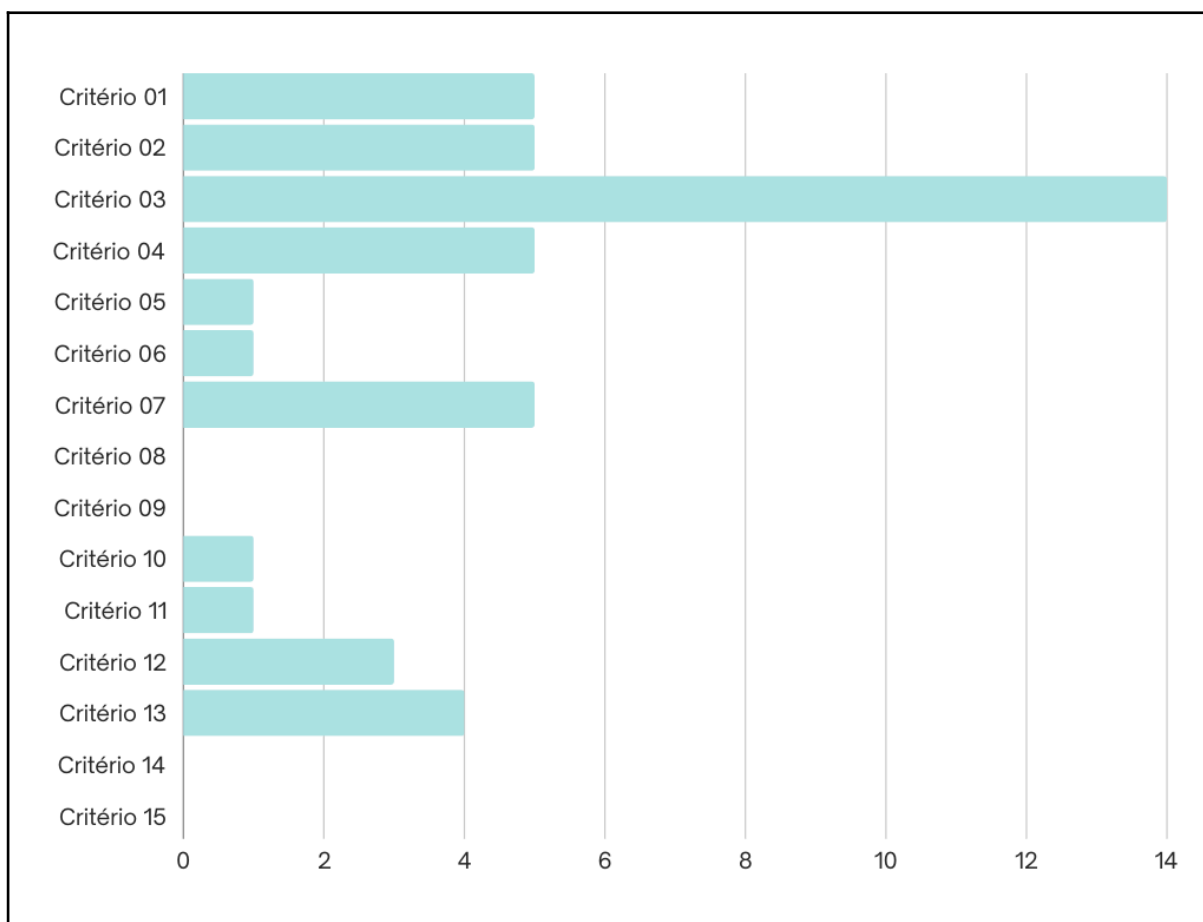
Fonte: elaboração própria (2024).

De acordo com o gráfico, os repositórios Aleia (Brasil) e DUnAs (Portugal) se destacam por definirem e disponibilizarem o maior número de critérios (5 critérios) para a atribuição de palavras-chave pelos autores científicos no processo de autoarquivamento. No entanto, ao considerar a quantidade possível de critérios identificados para essa finalidade, conforme descrito nesta pesquisa, observa-se que o problema da negligência na definição e disponibilização desses critérios é generalizado, afetando todos os repositórios e os países envolvidos neste estudo.

A partir do mapeamento ora apresentado (Quadro 9), constatou-se que nenhum dos 21 (vinte e um) repositórios analisados possui um documento formalmente denominado "Política" que inclua diretrizes específicas para a representação temática dos dados de pesquisa. Em teoria, a existência de tais documentos, de natureza primária, seria fundamental como base e referência para a criação de outros documentos secundários, como Manuais de Indexação, Guias de Indexação, Manuais de Autoarquivamento, Guias de Autoarquivamento e Tutoriais de Autoarquivamento. Na maioria dos repositórios, Guias, Manuais e Tutoriais são os únicos documentos acessíveis que, isoladamente, oferecem algum direcionamento, embora limitado, sobre a atribuição de palavras-chave aos dados de pesquisa a serem depositados.

Conforme detalhado no Apêndice A, 10 (dez) dos repositórios utilizam o projeto Dataverse como plataforma para a gestão de dados de pesquisa, compartilhando as mesmas instruções para grande parte dos seus processos. Estes repositórios são: Deposita Dados, Rede IBICT Cariniana Dataverse, Repositório de Dados de Pesquisa Unifesp, Arca Dados, Dados Abertos de Pesquisas, e REDU (Brasil); ioChem-BD (Espanha); e DUnAs, DadosIPB e DataRepositoriUM (Portugal). Nestes sistemas, ao buscar informações sobre o uso geral da plataforma, o depositante é direcionado a um mesmo Guia, que, embora forneça orientações sobre o processo de autoarquivamento, contém uma quantidade limitada de critérios para a atribuição de palavras-chave. Apenas um desses repositórios, o Deposita Dados, oferece um Guia adicional com mais diretrizes sobre o uso de palavras-chave.

A Figura 5 ilustra o nível de prevalência dos critérios analisados na pesquisa documental.

**Figura 5 - Nível de prevalência dos critérios analisados na pesquisa documental.**

Fonte: elaboração própria (2024).

Observou-se que o uso obrigatório de vocabulários controlados para a indexação nos repositórios analisados não é uma prática comum. Embora a maioria dos repositórios sugira a utilização desses vocabulários para o controle vocabular, poucos adotam estratégias que tornem seu uso obrigatório. Entre essas estratégias, pode-se citar a exigência de seleção obrigatória de termos por parte do depositante, a partir de uma lista pré-definida, elaborada com base em sistemas de organização do conhecimento existentes, nos termos mais pesquisados pelos usuários ou nos termos mais recorrentes nas produções científicas armazenadas no repositório.

O critério 03, que estabelece o "uso de vocabulário controlado a depender da escolha do depositante", foi o mais abordado nos documentos analisados, aparecendo em 14 (quatorze) dos 21 (vinte e um) repositórios investigados, o que corresponde a 66,67% do total. Esse dado indica uma preocupação relativa com o

controle vocabular na indexação realizada nos sistemas, mas ao transferir a responsabilidade pela escolha do vocabulário para o depositante os repositórios comprometem a consistência global da representação temática, o que pode resultar em uma recuperação ineficiente dos documentos. Para garantir uma representação coerente, seria mais eficaz que todos os depositantes utilizassem o mesmo vocabulário controlado, sob as mesmas diretrizes. No entanto, o critério 02, que recomenda o "uso de um vocabulário controlado específico", foi encontrado apenas em 5 (cinco) repositórios (23,81%).

O critério 01, que se refere à "definição própria para palavras-chave e/ou sua relevância para a representação de dados de pesquisa", apareceu em 5 (cinco) dos 21 (vinte e um) repositórios (23,81%). Esse critério é importante, pois ajuda o depositante a entender como o sistema de recuperação de informação define palavras-chave e o que se espera dele ao atribuir essas palavras ao conjunto de dados, alinhando sua representação temática às expectativas da instituição.

A "indicação da quantidade de termos/palavras-chave a serem incluídas" (critério 04) também foi encontrada em 5 (cinco) repositórios (23,81%). Esse critério é essencial, pois a quantidade de termos utilizados na indexação impacta diretamente na precisão e revocação da busca, influenciando a eficácia da recuperação da informação (Lancaster, 2004).

O critério 07, que trata da "orientação para o emprego de letras maiúsculas e/ou sinais de pontuação e/ou acentuação gráfica", foi identificado em 5 (cinco) repositórios (23,81%). Esse aspecto é particularmente importante em sistemas que fazem distinção entre maiúsculas e minúsculas, ou entre palavras pontuadas e não pontuadas.

"Lista de perguntas frequentes e/ou página de ajuda e/ou tutorial" (critério 12) são recursos adicionais recomendados por Terra *et al.* (2021) para serem disponibilizados aos depositantes a fim de esclarecerem dúvidas durante o processo de indexação. Nos repositórios analisados, apenas páginas de ajuda foram encontradas, em 3 (três) dos 21 (vinte e um) repositórios (14,29%). Contudo, essas

páginas não abordam questões relacionadas especificamente à representação temática de dados de pesquisa.

"Orientação para a ordenação de termos centrais e periféricos" (critério 05), um elemento central para a indexação, foi pouco abordada, sendo encontrada apenas em 1 (um) repositório (4,76%). Em alguns sistemas de recuperação da informação, a ordem das palavras-chave empregadas pode influenciar a recuperação da informação (Silva, 2022), mas essa é uma questão frequentemente negligenciada.

O critério 06, que trata do "controle de termos polissêmicos, sinônimos, diferenciação de termos homógrafos e/ou hierarquização de termos", relacionado aos vocabulários controlados, foi encontrado em apenas 1 (um) repositório (4,76%). Esses aspectos são fundamentais para garantir a consistência e a precisão na indexação de dados de pesquisa.

Embora a relação semântica entre termos seja fundamental na organização do conhecimento e deva ser refletida nas palavras-chave para garantir a efetividade da recuperação da informação, o critério 10, que trata da "orientação sobre a semanticidade entre termos empregados", foi identificado em apenas 1 (um) repositório (4,76%).

Os critérios "orientação para o emprego de termos singulares e plurais" (critério 08) e "orientação para a transcrição em língua estrangeira" (critério 09), que também são recomendados por Terra *et al.* (2021), não foram encontrados em nenhum dos documentos analisados. Esses critérios são importantes para manter a coerência e a consistência na atribuição de palavras-chave em diferentes idiomas e para garantir que a representação temática seja eficaz em contextos multilíngues. Em sua pesquisa, Terra *et al.* (2021) também investigou a implicação da repetição de palavras do título em palavras-chave (critério 14) e a repetição de palavras do resumo em palavras-chave (critério 15). Nos ambientes digitais específicos que foram analisados, não foi considerada proveitosa para a recuperação da informação a repetição de palavras do título em palavras-chave, embora esta tenha sido constatada como uma prática recorrente; por outro lado, por questões de

consolidação estratégica do desenvolvimento discursivo e semântico das palavras-chave, os pesquisadores consideraram importante a repetição de palavras dos resumos. No presente estudo, contudo, ambos critérios não estavam presentes em nenhum dos documentos analisados.

Por fim, tornou-se inviável realizar uma avaliação completa de dois critérios neste estudo. Na maioria dos repositórios, não foi possível obter informações claras sobre a ocorrência ou não da "validação dos termos por pessoa bibliotecária" (critério 13). Embora a ocorrência desse processo não seja explicitamente informado aos depositantes, é possível inferir que ele ocorre em alguns casos. De forma precisa, esse procedimento – fundamental para corrigir palavras-chave atribuídas inadequadamente – foi identificado em apenas 4 (quatro) dos 21 (vinte e um) repositórios analisados, o que corresponde a 19,05%. Em relação ao critério 11, "sugestão de termos no preenchimento de palavras-chave e/ou correção de erros ortográficos", embora tenha sido inicialmente considerado na análise, este foi descartado, posteriormente, da análise, já que a maioria dos repositórios não forneceu acesso às suas interfaces de *upload*, locais onde essas informações poderiam ser verificadas. A maioria dos repositórios requer adesão institucional para o depósito de dados ou adota um modelo fechado de *upload*. A constatação de que o sistema sugere termos ou corrige erros ortográficos foi possível em apenas 1 (um) dos 21 (vinte e um) repositórios (4,76%).

Em síntese, a análise das diretrizes constantes nos documentos de 21 (vinte e um) repositórios de dados de pesquisa estudados mostrou que eles não abordam todos os aspectos essenciais para a atribuição de palavras-chave pelos autores científicos, tendo base os critérios norteadores considerados neste estudo. Os repositórios oferecem instruções inconsistentes e superficiais para a indexação. Além disso, observa-se uma falta de união das informações, pois recomendações isoladas foram encontradas em vários documentos dentro de um mesmo repositório. Também se constatou que, na maioria dos casos, as diretrizes não são facilmente localizáveis nos repositórios, exigindo buscas extensas para encontrá-las.

Embora a atuação de profissionais bibliotecários nos repositórios estudados não tenha sido um objetivo específico desta pesquisa, é possível inferir que

repositórios institucionais, com maior frequência, prevêm a atuação desses profissionais na curadoria dos recursos informacionais que gerenciam, sob uma perspectiva mais estratégica, conferindo maior competitividade e eficiência na organização e na recuperação. Instituições produtoras de conhecimento científico, conscientes do papel de sua existência, que seguem uma perspectiva de disseminação otimizada do seu capital intelectual, são aptas a cultivarem uma filosofia de gerenciamento estratégico da produção gerada pela sua comunidade, incluindo em seu quadro profissional pessoas bibliotecárias capacitadas e ferramentas condicionadas à gestão dos materiais que zelam.

Avançando para a questão central desta seção, a pergunta que se impõe é: *"Como desenvolver diretrizes eficazes para a atribuição de palavras-chave a dados de pesquisa pelos autores científicos?"*

A resposta deve ser obtida, primeiro, mediante a reflexão estratégica das administrações institucionais responsáveis pelos respectivos repositórios digitais. Quanto mais especializada for a área temática associada a um conjunto de dados de pesquisa a ser autoarquivado, mais especializada será a natureza da indexação necessária, exigindo um conjunto de termos provenientes de um domínio do conhecimento técnico específico, compreendido principalmente por pesquisadores imersos em um campo mais restrito. Como consequência, isso demanda diretrizes mais profundas para guiar a atribuição de termos descritores e palavras-chave, além de uma infraestrutura tecnológica adequada que acomode esses termos.

A atuação de bibliotecários e outros profissionais da Ciência da Informação é crucial para aprimorar as condições da representação temática em repositórios digitais. Porém, essa atuação deve ser mais presente já na fase de implementação dos sistemas, momento em que frequentemente esses profissionais desempenham um papel secundário, cedendo espaço para especialistas em tecnologia e analistas de sistemas.

No entanto, esse cenário também representa uma oportunidade significativa. A familiaridade dos bibliotecários com recursos tecnológicos pode ser um

diferencial importante, sobretudo quando se considera que, como afirmam Marcondes e Sayão (2002, p. 45 ),

"os padrões de tecnologia da informação utilizados ou derivados da *Open Archives Initiative* têm um impacto potencial muito grande sobre os sistemas de informação em Ciência e Tecnologia, afetando substancialmente a maneira como bibliotecas e centros de documentação desempenham funções tradicionais como seleção, aquisição, registro/tratamento técnico, indexação, classificação e disseminação".

Quanto à elaboração de documentos objetivamente denominados *Políticas de Indexação*, o que ocorre, tendencialmente, é o desconhecimento generalizado por parte das instituições sobre a importância de um documento norteador que tenha como atributos transparência, completude e detalhamento maximizados para a finalidade de atribuição de palavras-chave a dados de pesquisa autoarquivados por autores científicos. No geral, como evidenciado por esta pesquisa, é compreendido pelas instituições que Guias e Manuais, sendo esses documentos mais econômicos em conteúdo, isoladamente cumprem o referido papel, porém, como constatado, em desacordo com a condução de uma representação temática eficiente. Contudo, é possível afirmar que, em algumas ocasiões, o Políticas de Indexação diretamente disponíveis ao usuário, sendo esses documentos tendencialmente mais extensos e densos em conteúdo, podem ser raramente utilizados por eles; indica-se seu uso, como ora defendido nesta pesquisa, como base para o desenvolvimento de outros documentos norteadores mais objetivos, quais sejam, Guias e Manuais.

Arrisca-se dizer que boa parte dos 15 (quinze) critérios norteadores para a atribuição de palavras-chave apresentados nesta pesquisa<sup>34</sup> são universais para a padronização dos procedimentos no referido processo. Em qualquer Área Temática,

---

<sup>34</sup> 1. definição própria para palavras-chave e/ou sua relevância para a representação de dados de pesquisa; 2. uso de um vocabulário controlado específico; 3. uso de vocabulário controlado a depender da escolha do depositante; 4. quantidade de termos/palavras-chave a serem incluída; 5. ordenação de termos centrais e periféricos; 6. controle de termos polissêmicos; sinônimos; diferenciação de termos homógrafos e/ou hierarquização de termos; 7. orientação para o emprego de letras maiúsculas e/ou sinais de pontuação e/ou acentuação gráfica; 8. orientação para o emprego de termos singulares e plurais; 9. orientação para a transcrição em língua estrangeira; 10. orientação sobre a semântica entre termos empregados; 11. sugestão de termos no preenchimento de palavras-chave e/ou correção de erros ortográficos; 12. disponibilização de Lista de perguntas frequentes e/ou Página de ajuda e/ou Tutorial; 13. validação dos termos por pessoa bibliotecária; 14. orientação sobre a repetição das palavras do título nas palavras-chave; e 15. orientação sobre a repetição das palavras do resumo nas palavras-chave.

esses critérios possuem a propriedade de guiar e lapidar a capacidade de julgamento do autor científico na representação temática de seus dados de pesquisa; e ainda, em um momento posterior, no caso validação dos termos por pessoa bibliotecária, otimizar o emprego dos termos e a recuperação informacional. Em documentos norteadores, um detalhamento suficiente para cada critério norteador apresentado, condicionado à natureza dos dados de pesquisa a que são direcionados, tende a estreitar a margem de dúvidas e a multiplicidade compreensiva por parte do seu público-alvo, ou seja, dos autores científicos.

Não menos importante, é necessária a consideração das mudanças tecnológicas que impactam nos sistemas de recuperação da informação científica, além das especificidades do funcionamento do *software* de cada repositório digital. Cada vez mais, esses sistemas estão condicionados a processarem a linguagem natural no contexto da indexação; questões como distorções relacionadas à pontuação, acentuação gráfica e ao emprego de letras maiúsculas/minúsculas em palavras-chave são facilmente contornadas por algoritmos. Na fase de recuperação informacional, esquemas de metadados como o *Dublin Core* já são capazes de recuperar os mesmos documentos com termos submetidos em maiúsculas ou minúsculas, com acentuação ou sem acentuação.

Por fim, o caso dos vocabulários controlados, amplamente indicados pela literatura científica para serem empregados na atribuição de palavras-chave, é necessária a compreensão sobre o nível intelectual do usuário que irá recuperar determinado documento armazenado, visto que, esses vocabulários tendem a ser, em maior ou menor grau, especializados terminologicamente; nem todo usuário que irá recuperar um documento em um repositório digital possui ampla familiarização com o domínio do conhecimento correspondente, fato que também pode demandar a utilização dessas ferramentas na etapa de recuperação da informação, para que distanciem esses usuários da linguagem natural. É necessária uma aprofundada avaliação sobre em quais casos o emprego de vocabulários controlados é vantajoso e em quais casos ele obstaculiza a recuperação de dados de pesquisa por toda ou por parte de uma comunidade usuária.

## 5 CONSIDERAÇÕES FINAIS

Nos repositórios digitais, a atribuição de metadados temáticos é fundamental para viabilizar a recuperação e reutilização de materiais digitais. Com a adoção de novas práticas, como o autoarquivamento de dados de pesquisa, surgem desafios adicionais para a representação temática, especialmente no que se refere a garantir que os termos atribuídos pelos autores científicos aos seus dados proporcionem informações suficientes sobre seu conteúdo, facilitando sua recuperação posterior. Nesse contexto, a criação de documentos com diretrizes claras para a representação temática torna-se uma solução importante, pois orienta a comunidade científica, auxiliando na redução de inconsistências no processo de indexação.

A literatura especializada da Ciência da Informação, focada na análise de Políticas de Indexação para o autoarquivamento de dados de pesquisa em repositórios digitais têm destacado problemas relacionados à qualidade dessas políticas e de suas diretrizes. Como apontado por Silva (2022), poucos repositórios de dados de pesquisa na Ibero-América possuem políticas de indexação formalizadas e fornecem orientações satisfatórias para a atribuição de palavras-chave pelos autores científicos. Em teoria, essas Políticas deveriam ser documentos norteadores, com atributos de transparência, completude e detalhamento, servindo como base para o desenvolvimento de documentos mais específicos, como Guias e Manuais.

Dada a possível falta de conhecimento por parte das instituições sobre a importância de Políticas de Indexação formalizadas, e a sua compreensão de que Guias e Manuais podem cumprir essa função, embora de forma inadequada, este estudo focou nas diretrizes isoladas encontradas em diversos documentos disponibilizados nos repositórios de dados de pesquisa de Brasil, Espanha e Portugal, registrados no diretório re3data.

A questão central da pesquisa foi: *Como as diretrizes de repositórios de dados de pesquisa vinculados ao Brasil, Espanha e Portugal orientam os autores científicos na atribuição de palavras-chave durante o processo de autoarquivamento?* O objetivo

final foi analisar essas diretrizes e fornecer subsídios para aprimorar as orientações existentes, criar novas diretrizes ou, idealmente, desenvolver Políticas de Indexação eficazes para sistemas de autoarquivamento de dados de pesquisa.

Para alcançar esse objetivo, a pesquisa foi de natureza aplicada e qualitativa, com abordagem descritiva e exploratória, utilizando procedimentos de coleta de dados bibliográficos e documentais. A análise detalhada de documentos disponibilizados nos repositórios de dados de pesquisa, como Manuais, Guias e Tutoriais, permitiu avaliar o estado atual da representação temática nos repositórios estudados.

Foi possível constatar que nenhum dos repositórios analisados possui um documento formalizado, especificamente denominado "Política", que aborde a representação temática dos dados de pesquisa. Embora esses documentos possam, em teoria, servir como a base para a criação de outros materiais orientadores, não foram encontrados esses documentos primários que fornecessem uma estrutura sólida e transparente para a atribuição de palavras-chave.

A partir da análise teórica e da revisão detalhada das diretrizes presentes nos 21 (vinte e um) repositórios de dados de pesquisa, foram identificados 15 (quinze) critérios para a atribuição de palavras-chave, que também serviram como critérios de análise. De forma geral, observou-se que as diretrizes encontradas não abordam todos os aspectos considerados essenciais, conforme a literatura especializada, para que os autores científicos conduzam uma representação temática eficaz de seus dados de pesquisa. Além disso, muitas diretrizes apresentavam pouca consistência e profundidade, e em alguns casos, informações isoladas estavam espalhadas em diferentes documentos dentro de um mesmo repositório. Na maioria dos casos, as diretrizes exigiram buscas exaustivas para serem localizadas.

No que diz respeito ao controle vocabular na indexação dos dados de pesquisa, observou-se que, embora haja alguma preocupação por parte das instituições, não são implementadas estratégias para tornar o uso de vocabulários controlados obrigatório. A maioria dos documentos recomenda o uso desses vocabulários, mas deixa a decisão ao critério do depositante. É importante ressaltar

que, para garantir uma representação temática mais consistente e eficaz, é fundamental que todos os depositantes utilizem o mesmo vocabulário controlado.

De acordo com os resultados desta pesquisa, a indicação de quantidade de termos a serem atribuídos aos dados de pesquisa; uma definição própria para palavras-chave e/ou sua relevância para a representação temática de dados de pesquisa; recomendação para o uso de um vocabulário controlado específico; e orientação para o emprego de letras maiúsculas e/ou sinais de pontuação e/ou acentuação gráfica, foram os aspectos ligeiramente abordados com mais frequência pelos documentos disponíveis. Outros aspectos importantes foram recorrentemente negligenciadas, tais como: orientação para o emprego de termos singulares e plurais; orientação sobre a semanticidade entre termos empregados; orientação para o controle de termos polissêmicos; sinônimos; diferenciação de termos homógrafos e/ou hierarquização de termos; orientação para a transcrição em língua estrangeira; orientação para a ordenação de termos centrais e periféricos; e disponibilização de Lista de perguntas frequentes e/ou Página de ajuda e/ou Tutorial. Nos documentos analisados não foram identificados critérios satisfatórios para a atribuição de palavras-chave; e os documentos direcionadores identificados não direcionam o autor científico da melhor maneira, visto que não refletem procedimentos estruturados e transparentes. O negligenciamento relativo à definição de cada um desses aspectos implica em prejuízos específicos para a recuperação dos dados de pesquisa depositados por autores científicos nos respectivos sistemas.

Conclui-se que a busca pela melhoria da qualidade das diretrizes para a indexação no autoarquivamento de dados de pesquisa deve ser uma prioridade para as instituições que gerenciam esses repositórios. Além disso, os critérios identificados neste estudo podem servir como base para o desenvolvimento de documentos mais consistentes e orientadores, que aprimorem a capacidade de julgamento dos autores científicos na atribuição de palavras-chave.

Dessa forma, mediante a análise documental conduzida, a questão de pesquisa definida para esta pesquisa foi respondida e o objetivo principal foi alcançado. A análise bibliográfica inicial, do mesmo modo, respondeu aos objetivos específicos aqui delimitados.

Como limitações deste estudo, que considerou o diretório re3data para a formação do universo de pesquisa, expõe-se a consequente não inclusão de outros repositórios de dados de pesquisa considerados importantes para estudo, por não estarem registrados neste diretório. Arrisca-se afirmar que uma investigação da mesma natureza desta pesquisa, conduzida com o uso de outros diretórios, permitiria a obtenção de resultados e conclusões distintas e valiosas.

Para pesquisas futuras com escopo similar e que utilizem procedimentos metodológicos semelhantes, recomenda-se alguma providência adequada quanto à possibilidade de acesso e análise das interfaces de *upload* de dados nos repositórios a serem considerados, visto que, na presente pesquisa, foram encontradas barreiras nesse sentido. Grande parte dos repositórios não permitiu acesso às referidas interfaces, locais onde, infere-se, informações relevantes para a execução da representação temática poderiam ser coletadas, tais como a existência de sugestões de termos no preenchimento de metadados e/ou correção instantânea de erros ortográficos. Tal impossibilidade de acesso ocorreu devido à obrigatoriedade de adesão institucional para o depósito de dados nesses repositórios ou então ao modelo fechado e restrito de *upload* empregados.

Por outra ótica, este estudo abre caminho para pesquisas futuras que explorem as perspectivas, necessidades e expectativas de diferentes agentes que utilizam as infraestruturas tecnológicas analisadas, como bibliotecários e autores científicos que arquivam dados de pesquisa. Seria importante a compreensão do seu posicionamento em relação às condições oferecidas pelos respectivos sistemas de recuperação para que a representação temática dos dados de pesquisa seja efetivamente realizada. Além disso, poderia ser interessante investigar quais áreas temáticas geram maior interesse e preocupação dentro desse contexto.

Enfim, a busca pela qualidade de documentos norteadores na perspectiva estudada deve, permanentemente, fazer parte dos objetivos de cada instituição mantenedora. Após a formulação ou a adequação das diretrizes para a indexação, orienta-se que o documento criado ou editado seja disponibilizado na interface principal do respectivo sistema de recuperação de informação, de forma facilmente localizável. Para as nações envolvidas nesta pesquisa, Brasil, Espanha e Portugal,

infere-se que há a necessidade de uma cooperação mais próxima no âmbito da representação temática de dados de pesquisa, visto que, colaborações acadêmicas e científicas são fundamentais para o avanço e desenvolvimento da Ciência, para a sua divulgação e para a transferência de conhecimentos.

## REFERÊNCIAS

ALBAGLI, S.; CLINIO, A.; RAYCHTOCK, S. Ciência aberta: correntes interpretativas e tipos de ação. **Liinc**, Rio de Janeiro, RJ, v.10, n.2, p. 434-450, nov. 2014. Disponível em: <https://periodicos.ufpb.br/index.php/pbcib/article/view/23180>. Acesso em: 18 fev. 2023.

ANDRADE, M. A. L.; SABBAG, D. M. A. Política de indexação: um estudo nas Bibliotecas Públicas do Estado do Rio de Janeiro. **Biblioteca Escolar em Revista**, Ribeirão Preto, v. 5, n. 2, p. 75-97, 2017.

AQUINO, I. S.; AQUINO, I. S. Análise sobre a forma da escrita de palavras-chave em artigos científicos na área de ciências agrárias publicados no período de 1999 a 2011. **Encontros Bibli**: revista eletrônica de Biblioteconomia e Ciência da Informação, Florianópolis, v. 18, n. 37, p. 227-238, maio/ago. 2013.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **NBR 12676**: Métodos para análise de documentos - determinação de seus assuntos e seleção de termos de indexação. Rio de Janeiro, 1992.

AVENTURIER, P. **Princípios FAIR**: critérios de qualidade para dados de pesquisa. 2017. Disponível em: <https://publicient.hypotheses.org/1456>. Acesso em: 06 jul. 2023.

BARDIN, L. **Análise de Conteúdo**. Lisboa: Edições 70, 1977.

BELLI, S.; DÍEZ, R.; RAMOS, M. L.; RESINA J.; DELGADO, M. R. **Desafios para as políticas de ciência aberta na Ibero-América**: redes, repositórios e multilinguismo, 2023. Disponível em: <https://oei.int/pt/publicacoes/desafios-para-las-politicas-de-ciencia-abierta-en-iberoamerica-redes-repositorios-y-multilinguismo>. Acesso em: 16 jun. 2024.

**BERLIN DECLARATION ON OPEN ACCESS TO KNOWLEDGE IN THE SCIENCES AND HUMANITIES**. 2003. Disponível em: <https://openaccess.mpg.de/Berlin-Declaration>. Acesso em: 19 nov. 2023.

BERNERS-LEE, T. **Linked data World Wide Web Consortium**, 2006. Disponível em: <http://www.w3.org/DesignIssues/LinkedData.html>. Acesso em: 13 abr. 2024.

BERNERS-LEE, T. **Linked Data**. [s. l.], 2010. Disponível em: <https://www.w3.org/DesignIssues/LinkedData.html>. Acesso em: 13 abr. 2024.

BOCCATO, V. R. C.; TORQUETTI, M. C. Interoperabilidade entre linguagens de indexação como recurso de modelagem de repertório terminológico de coordenadorias de comunicação social em ambientes universitários: uma proposta metodológica. **Informação & Informação**, v. 17, n. 3, p. 76-101, 2012. Disponível em:

<http://www.uel.br/revistas/uel/index.php/informacao/article/view/10800/pdf>. Acesso em: 2 mar. 2023.

BRANDAU, R.; MONTEIRO, R.; BRAILE, D. M. Importância do uso correto dos descritores nos artigos científicos. **Revista Brasileira de Cirurgia Cardiovascular**, São José do Rio Preto, v. 20, n. 1, p. VII-IX, jan./mar. 2005.

BUDAPEST OPEN ACCESS INITIATIVE. **Iniciativa de Budapeste pelo acesso aberto**. 2002. Disponível em:

<https://www.budapestopenaccessinitiative.org/translations/portuguesetranslationsinitiative.org/boai10/portuguese-brazilian-translation/>. Acesso em: 14 jun. 2023.

BUENO DE LA FUENTE, G.; AGUSTÍN-LACRUZ, G. B. F.; FUJITA, M. S. L.; TERRA, A. L. Knowledge organisation in institutional repositories: a case study on policies and procedures manuals in the Ibero-American environment. **The Electronic Library**, Vol. ahead-of-print No. ahead-of-print. 2023. Disponível em:

[https://www.emerald.com/insight/content/doi/10.1108/EL-05-2023-0128/full/html?utm\\_source=rss&utm\\_medium=feed&utm\\_campaign=rss\\_journalLatest](https://www.emerald.com/insight/content/doi/10.1108/EL-05-2023-0128/full/html?utm_source=rss&utm_medium=feed&utm_campaign=rss_journalLatest). Acesso em: 15 jul. 2023.

CARNEIRO, M. V. Diretrizes para uma política de indexação. **Revista da Escola de Biblioteconomia da UFMG**, Belo Horizonte, v. 14, n. 2, p. 221-241, set. 1985.

CASARIN, H. C. S.; CASARIN, S. J. **Pesquisa científica: da teoria à prática**. Curitiba: Intersaberes, 2012. 200 p.

CESARINO, M. A. N. Sistemas de recuperação da informação. **Revista da Escola de Biblioteconomia da UFMG**, Belo Horizonte, n. 14, v. 2, p. 157-168, set. 1985.

Disponível em: <http://www.brapci.inf.br/index.php/article/download/13794>. Acesso em: 1 jun. 2023.

COMISSÃO EUROPEIA. **Guidelines to the rules on open access to scientific publications and open access to research data in horizon 2020**. 2017. Disponível em:

[https://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/oa\\_pilot/h2020-hi-oa-pilotguide\\_en.pdf](https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilotguide_en.pdf). Acesso em: 13 jun. 2024.

CONARQ, CONSELHO NACIONAL DE ARQUIVOS. **Diretrizes para a implementação de Repositórios Arquivísticos Digitais Confiáveis-RDC-ARQ**. Rio de Janeiro: CONARQ, 2022. Disponível em: <https://bit.ly/3nJwSyd>. Acesso em: 31 ago 2024.

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: 30 jul. 2021.

CREASER, C. Scholarly communication and access to research output. *In*: EVANS, WENDY, BAKER, DAVIS. **Libraries and Society**: role, responsibility and future in an age of change. EBSCO Publishing. 2011. p. 53-66.

CURTY, R. G. *et al.* Attitudes and norms affecting scientists' data reuse. **PLOS ONE**, [s. l.], v. 12, n. 12, p. e0189288, 2017.

FATTAHI, R. Library cataloging and abstracting and indexing services: reconciliation of principles in the online environment? **Library Review**, Bradford, v. 47, n. 4, p. 211- , 1998.

FREITAS, M. P.; DAL'EVEDOVE P. R.; TARTAROTTI, R. C. D. Políticas de autoarquivamento em repositórios institucionais brasileiros: estudo analítico do metadado assunto. **Páginas a&b**: arquivos e bibliotecas, 169–175. 2021. Disponível em: <https://ojs.letras.up.pt/index.php/paginasaeb/article/view/10245>. Acesso em: 23 fev. 2023.

FUJITA, M. S. L. A identificação de conceitos no processo de análise de assunto para indexação. **Revista Digital de Biblioteconomia e Ciência da Informação**, Campinas, v. 1, n. 1, p. 60-90, jul./dez. 2003.

FUJITA, M. S. L. A representação documentária de artigos científicos em educação especial: orientação aos autores para determinação de palavras chaves. **Rev. bras. educ. espec.**, Marília, v. 10, n. 03, p. 257-272, 2004. Disponível em: [http://educa.fcc.org.br/scielo.php?script=sci\\_arttext&pid=S1413-65382004000300002&lng=pt&nrm=iso](http://educa.fcc.org.br/scielo.php?script=sci_arttext&pid=S1413-65382004000300002&lng=pt&nrm=iso). Acesso em: 30 jul. 2023.

FUJITA, M. S. L.; SANTOS, L. B. P. D. Política de indexação em bibliotecas universitárias: estudo diagnóstico e analítico com pesquisa participante. **Transinformação**, v. 28, n. 1, p. 59-76, 2016. Acesso em: 04 jun. 2023.

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. Acesso em: 30 jul. 2023.

GERHARDT, T. E.; SILVEIRA, D. T. (Org.). **Métodos de Pesquisa**. 1ª Ed. Porto Alegre: Editora da UFRGS, 2009.

GIL, A. C. **Métodos e técnicas de pesquisa social**. 6. ed. São Paulo: Atlas, 2011.

GOLUB, K.; GÖRANSSON, E.; FOKA, A.; HUVILA, I. **Digital humanities in Sweden and its infrastructure**: Status quo and the sine qua non. *Digital Scholarship in the Humanities*, 2020. 35(3), 547-556. Disponível em: <https://doi.org/10.1093/llc/fqz042>. Acesso em: 20 jun. 2024.

GOMES, R. F.; LIMA, G. N. B. O. **Importância da política de indexação para as unidades de informação: uma revisão sistemática da literatura**. , v. 27, n. online, n. 1, p. 210-236, 2021. Acesso em: 22 out. 2023.

GONÇALVES, A. L. Uso de resumos e palavras-chave em Ciências Sociais: uma avaliação. **Encontros Bibli: revista eletrônica de Biblioteconomia e Ciência da Informação**, Florianópolis, v. 13, n. 26, p. 78-93, out. 2008. Disponível em: <https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2008v13n26p78>. Acesso em: 9 jul. 2023.

GOV BR. **O que é Governo Aberto**, s/d. Disponível em: <https://www.gov.br/cgu/pt-br/governo-aberto/governo-aberto-no-brasil/principios#:~:text=Governo%20Aberto%20%C3%A9%20uma%20cultura,solucionar%20quest%C3%B5es%20de%20interesse%20p%C3%ABlico>. Acesso em: 13 jun. 2024.

GULARTE, A. A. **A estrutura europeia para Dados de Pesquisa em Ciências Sociais: reflexões e perspectivas**. 2021. Dissertação (Mestrado em Ciência da Informação) – Universidade Federal do Rio Grande do Sul, Porto Alegre. Disponível em: <https://lume.ufrgs.br/handle/10183/233316>. Acesso em: 19 nov. 2023.

HEFCE *et al.* **Concordat on Open Research Data**. [s. l.: s. n.], 2016. Disponível em: <https://www.ukri.org/wp-content/uploads/2020/10/UKRI-020920-ConcordatonOpenResearchData.pdf>. Acesso em: 13 abr. 2024.

HENNING, P. C.; RIBEIRO, C. J. S.; SANTOS, L. O. B.; SANTOS, P. X. 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, 2019. Disponível em: <https://brapci.inf.br/index.php/res/v/113770>. Acesso em: 10 mar. 2023.

HJØRLAND, B. Concept theory. **JASIST**, v. 60, n. 8, p. 1519-1536, 2009.

HJØRLAND, B. Is classification necessary after Google? **ISKO**. 2012. p. 19-30.

HODSON, S. JONES, S.; COLLINS, S. GENOVA, F.; HARROWER, N.; MIETCHEN, D.; PETRAUSKAITÉ, R.; WITTENBURG, P. **FAIR Data Action Plan: interim recommendations and actions from the European Commission Expert Group on FAIR data**. 2018. June. 21 p. Disponível em: <https://zenodo.org/record/1285290#.WyOWGSB9jt4>. Acesso em: 6 jul. 2023.

IBEKWE-SANJUAN, F.; BOWKER, G. C. Implications of big data for knowledge organization. **Knowledge Organization**, v. 44, n. 3, p. 187-198. 2017.

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: 23 jul. 2023.

INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA. **Repositórios digitais**. [S. l.], 2012. Disponível em: <http://sitehistorico.ibict.br/informacao-para-ciencia-tecnologia-e-inovacao%20/repositorios-digitais/repositorios-brasileiros>. Acesso em: 17 fev. 2023.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. **ISO 25964-1:2011** Information and documentation – Thesauri and interoperability with other vocabularies – Part 1: Thesauri for information retrieval. Geneva: International Organization for Standardization, 2011.

ISOTANI, S.; PINTO, I. I. B. S. **Dados abertos conectados**. São Paulo: Novatec, 2015. E-book. Disponível em: <https://search.ebscohost.com/login.aspx?direct=true&db=cat07377a&AN=sabi.000985290&site=eds-live>. Acesso em: 13 abr. 2024.

KEEFER, A. **Encuesta sobre publicación científica y auto-archivo**: resumen de los resultados de la encuesta de opinión. Madrid: Universidad Complutense de Madrid, 2007. Documento de trabajo. Disponível em: . Acesso em: 1 jul. 2023.

KITCHIN, R. **The Data Revolution**: Big Data, Open Data, Data Infrastructures & Their Consequences. London: SAGE Publications Ltd, 2014. E-book. Disponível em: <http://methods.sagepub.com/book/the-data-revolution>. Acesso em: 19 nov. 2023.

KOBASHI, N. Y. Análise documentária e representação da informação. **Revista Informare**, Rio de Janeiro, v. 2, n. 2, p. 5-27, 1996.

KURAMOTO, H. Informação científica: proposta de um novo modelo para o Brasil. **Ciência da Informação**, Brasília, v. 35, n. 2, p. 91-102, maio/ago. 2006. Disponível em: <http://www.scielo.br/pdf/ci/v35n2/a10v35n2.pdf>. Acesso em: 22 fev. 2023.

LAKATOS, E. M.; MARCONI, M. de A. **Fundamentos de metodologia científica**. 6. ed. 5. reimp. São Paulo: Atlas, 2007.

LANCASTER, F. W. **Indexação e resumos**: teoria e prática 2. ed. rev. atual. Brasília: Briquet de Lemos/Livros, 2004.

LEITE, F. C. L. **Como gerenciar e ampliar a visibilidade da produção científica brasileira**: repositórios institucionais de acesso aberto. Brasília: Ibict, 2009. Disponível em: <http://livroaberto.ibict.br/bitstream/1/775/4/Como%20gerenciar%20e%20ampliar%20a%20visibilidade%20da%20informa%C3%A7%C3%A3o%20cient%C3%AAdfica%20brasileira.pdf>. Acesso em: 14 jul. 2023.

LIMA, J. L. O.; ALVARES, L. **Organização e representação da informação e do conhecimento**. In: ALVARES, L. (org.). Organização da informação e do conhecimento: conceitos, subsídios interdisciplinares e aplicações. São Paulo: B4 Editores, 2012. p. 21-48.

LYNCH, C. A. Big data: how do your data grow? **Nature**, v. 455, n. 7209, p. 28–29, 2008. Disponível em: <http://www.nature.com/nature/journal/v455/n7209/full/455028a.html?foxtrotcallback=true>. Acesso em: 6 jul. 2023.

MARCONDES, C. H.; SAYÃO, L. F. Documentos digitais e novas formas de cooperação entre sistemas de informação em C&T. **Ciência da Informação**, Brasília, v. 31, n. 3, set/dez. 2002.

MEDEIROS, G. M. **Organização da informação em repositórios digitais**: implicações do auto-arquivamento na representação da informação. Dissertação (Mestrado em Ciência da Informação) – Programa de Pós-Graduação em Ciência da Informação, Universidade Federal de Santa Catarina, Florianópolis, 2010

MELERO, R. **Políticas sobre el libre acceso a la producción científica y la respuesta de los autores**. In: CONGRESO DE COMUNICACIÓN DE LA CIENCIA, 4., Madrid, 2007. Disponível em: <http://eprints.rclis.org/10429>. Acesso em: 1 jul. 2023.

MIGUÉIS, A.; NEVES, B.; SILVA A. L.; TRINDADE A.; BERNARDES J. A. 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**. 4:2 (2013) 112-125. 2013. Disponível em: <https://www.revistas.usp.br/incid/article/view/69284/71742>. Acesso em: 20 fev. 2023.

MONTEIRO, E. C. S. A.; AFFONSO, E. P.; BORBA, V. U.; SANTANA, R. C. G. A privacidade e os planos de gerenciamento de dados de repositórios de dados científicos. **Informação & Tecnologia**, v. 4, n. 1, p. 35-53, 2017. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/100543>. Acesso em: 27 jan. 2022.

NATIONAL INFORMATION STANDARDS ORGANIZATION. **Understanding Metadata**. Bethesda, MD: NISO Press, 2004. Disponível em: <http://www.niso.org/standards/resources/UnderstandingMetadata.pdf>. Acesso em: 20 fev. 2023.

NSF. **National Science Foundation**, 2020. Disponível em: <https://www.nsf.gov/>. Acesso em: 6 jul. 2023.

OLIVEIRA, L. P.; FUJITA, M. S. L.; DAL'EVEDOVE, P. R.; MARTÍNEZ-ÁVILA, D. 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**, 25(4), 2021. Disponível em: <https://periodicos.ufmg.br/index.php/pci/article/view/26936>. Acesso em: 18 fev. 2023.

OPEN GOV DATA. **The Annotated 8 Principles of Open Government Data**, 2007. Disponível em: <https://opengovdata.org/>. Acesso em: 13 jun. 2024.

**OPEN SCIENCE: RESEARCH DATA**. 2024. Disponível em: <https://www.ouvrirlascience.fr/wp-content/uploads/2024/03/24-02-28-Donnees-EN-WEB.pdf>. Acesso em: 14 jun. 2024.

ORGANIZATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT. **OECD Principles and Guidelines for Access to Research Data from Public Funding**, 2004.

Disponível em:

[google.com/search?q=oeed+&sca\\_esv=cff19ea8ba3a868e&sca\\_upv=1&rlz=1C1GCE A\\_enBR981BR981&sxsrf=ADLYWIK7agn7dY-PiH36CfnYLMbCQiEHeA%3A1718311303947&ei=h1lrZvjoMNPZ1sQPsaSbqA4&udm=&ved=0ahUKEwi4ylG3uNmGAXTrJUC HTHSBuUQ4dUDCBA&uact=5&oq=oeed+&gs\\_lp=Egxnd3Mtd2l6LXNlcnAiBW9lY2QgSlgBUABYAHAAeACQAQCYAQCgAQCqAQC4AQPIAQD4AQGYAgCgAgCYAwCSBwCgBwA&scient=gws-wiz-serp](https://www.google.com/search?q=oeed+&sca_esv=cff19ea8ba3a868e&sca_upv=1&rlz=1C1GCE A_enBR981BR981&sxsrf=ADLYWIK7agn7dY-PiH36CfnYLMbCQiEHeA%3A1718311303947&ei=h1lrZvjoMNPZ1sQPsaSbqA4&udm=&ved=0ahUKEwi4ylG3uNmGAXTrJUC HTHSBuUQ4dUDCBA&uact=5&oq=oeed+&gs_lp=Egxnd3Mtd2l6LXNlcnAiBW9lY2QgSlgBUABYAHAAeACQAQCYAQCgAQCqAQC4AQPIAQD4AQGYAgCgAgCYAwCSBwCgBwA&scient=gws-wiz-serp). Acesso em: 13 jun. 2024.

ORGANIZATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT. **OECD Principles and Guidelines for Access to Research Data from Public Funding**, 2007.

Disponível em:

[https://www.oecd-ilibrary.org/science-and-technology/oeed-principles-and-guidelines-for-access-to-research-data-from-public-funding\\_9789264034020-en-fr](https://www.oecd-ilibrary.org/science-and-technology/oeed-principles-and-guidelines-for-access-to-research-data-from-public-funding_9789264034020-en-fr). Acesso em: 13 jun. 2024.

PAMPEL, H.; VIERKANT, P.; SCHOLZE, F.; BERTELMANN, R.; KINDLING, M.; KLUMP, J.; GOEBELBECKER, H.; GUNDLACH, J.; SCHIRMBACHER, P.; DIEROLF, U. Making research data repositories visible: the re3data.org Registry. **PLoS One**, San Francisco, v. 8, n.11, 2013. Disponível em: . Acesso em: 17 out. 2023.

PAVÃO, C. M. G.; PASSOS, P. C. S. J.; GABRIEL JUNIOR, R. F.; VANZ, S. A. S.; BORGES, E. N.; AZAMBUJA, L. A. B. **Acesso aberto a dados de pesquisa no Brasil: repositórios brasileiros de dados de pesquisa: relatório 2018**. Disponível em: <http://hdl.handle.net/10183/185138>. Acesso em: 09 set. 2023.

PERIOTTO, C.; SARVO, D. O.; LOZANO, M. C. **Manual de Autodepósito de Dados de Pesquisa no RI UFSCar**. 2023. Disponível em:

<https://repositorio.ufscar.br/static/ri-ufscar-manual-autodeposito-dados.pdf>. Acesso em: 13 jun. 2024.

PINHEIRO, L. V. R. Medidas de consistência da indexação: interconsistência. **Ciência da Informação**, Rio de Janeiro, v.7, n.2, p.109-114, 1978.

PUBLIC RESOURCE. **Welcome to Public.Resource.Org!** Disponível em:

<https://public.resource.org/>. Acesso em: 13 jun. 2024.

RAUEN, C. V. A relevância de uma política nacional de acesso aberto a dados de pesquisa. **Revista Construção**, jan. 2018. Disponível em:

<http://revistaconstrucao.org/ciencia-e-tecnologia/relevancia-de-umapoliticanacional-de-acesso-aberto-dados-de-pesquisa>. Acesso em: 31 maio 2023.

RIBES, D. **O núcleo de uma infraestrutura de pesquisa**. In FUSSEL, S.; LUTTES, W.; MORRIS, M. R.; REDDY, M. (Eds.), *Anais da 17ª conferência ACM sobre trabalho cooperativo apoiado por computador e computação social*, 2014. Baltimore (Maryland), Estados Unidos, 15 a 19 de fevereiro de 2014 (págs. 574-587).

Associação de Máquinas de Computação. Disponível em:  
<https://doi.org/10.1145/2531602.2531700>. Acesso em: 20 jun. 2024.

RICE, R.; SOUTHALL, J. **The data librarian's handbook**. London: Facet Publishing, 2016.

RUBI, M. P.; FUJITA, M. S. L. Elementos de política de indexação em manuais de indexação de sistemas de informação especializados. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 8, n. 1, p. 66-77, jan./jun. 2003.

RUBI, M. P.; FUJITA, M. S. L. **O ensino de procedimentos de política de indexação na perspectiva do conhecimento organizacional**: uma proposta de programa para a educação à distância do bibliotecário. 2006. Disponível em:  
<http://hdl.handle.net/20.500.11959/brapci/32596>. Acesso em: 05 nov. 2023.

RUBI, M. P. **Política de indexação para construção de catálogos coletivos em bibliotecas universitárias**. Marília, 2008. 169f. Tese (Doutorado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista Júlio de Mesquita Filho.

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. 2023.

SANT'ANA, R. C. G. **Ciclo de vida dos dados**: Uma perspectiva a partir da Ciência da Informação. *Informação e Informação*, Londrina-PR, v.21, n.2, 2016. p.116-142. Disponível em:  
<http://www.uel.br/revistas/uel/index.php/informacao/article/view/27940/20124>. Acesso em: 29 ago. 2024.

SAYÃO, L. F.; SALES, L. F. **Guia de gestão de dados de pesquisa para bibliotecários e pesquisadores**. Rio de Janeiro: CNEN/IEN, 2015. E-book. Disponível em:  
[https://www.researchgate.net/publication/286455028\\_Guia\\_de\\_gestao\\_de\\_dados\\_de\\_pesquisa\\_para\\_bibliotecarios\\_e\\_pesquisadores](https://www.researchgate.net/publication/286455028_Guia_de_gestao_de_dados_de_pesquisa_para_bibliotecarios_e_pesquisadores). Acesso em: 19 nov. 2023.

SAYÃO, L. F.; SALES, L. F. Algumas considerações sobre os repositórios digitais de dados de pesquisa. **Informação & Informação**, v. 21, n. 2, p. 90- 115, 2016.

SCIENTIFIC ELECTRONIC LIBRARY ONLINE (SciELO). **Guia de preparação de dados de pesquisa** [online]. 2021. Disponível em:  
[https://wp.scielo.org/wpcontent/uploads/Guia\\_preparacao\\_pt.pdf](https://wp.scielo.org/wpcontent/uploads/Guia_preparacao_pt.pdf). Acesso em: 13 jun. 2024.

SCIENTIFIC ELECTRONIC LIBRARY ONLINE (SciELO). **Critérios, política e procedimentos para a admissão e a permanência de periódicos na Coleção SciELO Brasil** [online]. 2022. Disponível em:  
<https://www.scielo.br/media/files/20220900-criterios-scielo-brasil.pdf>. Acesso em: 1 set 2024.

SCIENTIFIC ELECTRONIC LIBRARY ONLINE (SciELO). **Guia de curadoria de dados de pesquisa para equipes editoriais** [online]. 2023. Acesso em: 1 set 2024.

SENDRA, A.; LATE, E.; KUMPULAINEN, S. More than data repositories: perceived information needs for the development of social sciences and humanities research infrastructures. 2023. **Information Research an international electronic journal** 28(4):83-10. Disponível em: [https://www.researchgate.net/publication/376598348\\_More\\_than\\_data\\_repositories\\_perceived\\_information\\_needs\\_for\\_the\\_development\\_of\\_social\\_sciences\\_and\\_humanities\\_research\\_infrastructures](https://www.researchgate.net/publication/376598348_More_than_data_repositories_perceived_information_needs_for_the_development_of_social_sciences_and_humanities_research_infrastructures). Acesso em: 20 jun. 2024.

SILVA, F.C.C.; SILVEIRA, L. O ecossistema da Ciência Aberta. **Transinformação**, v.31, e190001, 2019. Disponível em: <https://www.scielo.br/j/tinf/a/dJ89vRg94Qxtf6Y7M49Hztr/#>. Acesso em: 06 abr. 2024.

SILVA, W. K. P. **Atribuição de palavras-chave na prática do arquivamento**: estudo das diretrizes de repositórios de dados de pesquisa ibero-americanos. 2022. Dissertação (Mestrado em Ciência da Informação) – Universidade Federal de São Carlos, São Carlos, 2022. Disponível em: [https://repositorio.ufscar.br/bitstream/handle/ufscar/17011/Disserta%C3%A7%C3%A3o\\_WendySilva.pdf?sequence=2&isAllowed=y](https://repositorio.ufscar.br/bitstream/handle/ufscar/17011/Disserta%C3%A7%C3%A3o_WendySilva.pdf?sequence=2&isAllowed=y). Acesso em: 20 fev. 2023.

SILVA, F.C.C. Movimento open data em bibliotecas: desafios e oportunidades. *In*: Seminário do Grupo de Pesquisa MHTX: perspectivas em representação e organização do conhecimento: atualidades e tendências na relação universidade-empresa, 3., Belo Horizonte, Brasil. **Anais [...]** Belo Horizonte: ECI/UFMG, 2018.

STREHL, L. Avaliação da consistência da indexação realizada em uma biblioteca universitária de artes. 1998. **Ciência Da Informação**, 27(3). Disponível em: <https://revista.ibict.br/ciinf/article/view/787>. Acesso em: 20 nov. 2023.

SUBER, P. The primacy of authors in achieving open access. **Nature**, 2004. Disponível em: <https://www.nature.com/nature/focus/accessdebte/24.html>. Acesso em: 04 abr. 2013.

TARTAROTTI, R. C. D.; DAL'EVEDOVE, P. R.; FUJITA, M. S. L. Biblioteconomia de dados em repositórios de pesquisa: perspectivas para a atuação bibliotecária. **Informação & Informação**, v. 24, n. 3, p. 207-226, 2019. Disponível em: <https://brapci.inf.br/index.php/res/v/134221>. Acesso em: 24 fev. 2023.

TARTAROTTI, R. C. D.; DAL'EVEDOVE, P. R.; FUJITA, M. S. L. Indexação em repositórios digitais: uma incursão na produção científica brasileira de Ciência da Informação. **Scire**. 26:2 (jul.-dic. 2020) 13-18. Disponível em: [https://www.researchgate.net/publication/347521600\\_Indexacao\\_em\\_repositorios\\_d](https://www.researchgate.net/publication/347521600_Indexacao_em_repositorios_d)

igitais\_uma\_incursao\_na\_producao\_cientifica\_brasileira\_de\_Ciencia\_da\_Informacao. Acesso em: 20 fev. 2023.

TERRA, A. L.; AGUSTIN-LACRUZ, C.; BERNARDES, Ó.; FUJITA, M. S. L.; BUENO DE LA FUENTE, G. 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**, 47, 102268. 2021. Disponível em [https://www.researchgate.net/publication/346668214\\_Subject-access\\_metadata\\_on\\_ETD\\_supplied\\_by\\_authors\\_A\\_case\\_study\\_about\\_keywords\\_titles\\_and\\_abstracts\\_in\\_a\\_Brazilian\\_academic\\_repository](https://www.researchgate.net/publication/346668214_Subject-access_metadata_on_ETD_supplied_by_authors_A_case_study_about_keywords_titles_and_abstracts_in_a_Brazilian_academic_repository) . Acesso em: 20 nov. 2023.

TORINO, E. Políticas em Repositórios Digitais: das Diretrizes à Implementação, 2017, p. 91-114. In: VECHIATO F.; GUEDES, C. KOSHIYAMA, D.; MOURA, E.; TORINO, E.; MAIA, M. A.; MARQUES, T. (org.) **Repositórios digitais: teoria e prática**, Curitiba. 2017. Disponível em: <https://repositorio.ufrn.br/handle/123456789/24189>. Acesso em: 14 jun. 2024.

UNESCO. **Draft text of the UNESCO Recommendation on Open Science**. In: Intergovernmental Meeting of Experts (Category II), online, 2021. Disponível em: <https://unesdoc.unesco.org/ark:/48223/pf0000376893> Acesso em: 07 abr. 2024.

VEIGA, V.; MACENA, L. G. O autoarquivamento nos repositórios institucionais brasileiros: um estudo exploratório. **PontodeAcesso**, Salvador, v. 9, n. 3, p. 35-47, dez. 2015. Disponível em: <https://portalseer.ufba.br/index.php/revistaici/article/view/15107>. Acesso em: 20 fev. 2023.

VIANA, J. M. D. A. **Representação colaborativa de dados científicos**: estudo na rede de repositórios de dados científicos do estado de São Paulo. 2020. Dissertação (Mestrado em Ciência da Informação) – Universidade Federal de São Carlos, São Carlos, 2020. Disponível em: <https://repositorio.ufscar.br/handle/ufscar/13027>. Acesso em: 20 fev. 2023.

VIANA, J. M. D. 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&B, Arquivos e Bibliotecas** (Portugal), n. Especial, p. 192-196, 2021. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/157351>. Acesso em: 24 fev. 2023.

VIANA, C. L. M; MÁRDERO-ARELLANO, M. A.; SHINTAKU, M. **Repositórios institucionais em Ciência e Tecnologia**: uma experiência de customização do Dspace. In: PROCEEDINGS SIMPÓSIO INTERNACIONAL DE BIBLIOTECAS DIGITAIS, 3., 2005, São Paulo. Anais [...] São Paulo. Disponível em: <http://eprints.rclis.org/7168/1/viana358.pdf>. Acesso em: 20 fev. 2023.

WEITZEL, S. R. O papel dos repositórios institucionais e temáticos na estrutura da produção científica. **Em questão**, v. 12, n. 1, p. 51-71, 2006. Disponível em: <https://seer.ufrgs.br/EmQuestao/article/view/19/7>. Acesso em: 20 fev. 2023.


WILKINSON, M. D. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. **Scientific Data**, [s.l.], v. 3, 160018, 2016. Disponível em: <https://www.nature.com/articles/sdata201618>. Acesso em: 19 nov. 2023.

WORLD HEALTH ORGANIZATION (WHO). **A coordinated global research roadmap: 2019 novel coronavirus**. Geneva, Switzerland: The World Health Organization. 2020. Disponível em: <https://www.who.int/publications/m/item/a-coordinated-global-research-roadmap>. Acesso em: 30 mar. 2024.

XIA, J. Positioning Open Access Journals in a LIS Journal Ranking. **College & Research Libraries**, Chicago, v. 73, n. 2, p. 134-145, Mar. 2012.

ZÁRATE, M. *et al.* Improving Open Science Using Linked Open Data: CONICET Digital Use Case. **Journal of Computer Science & Technology**, La Plata, v. 19, n. 1, p. 45-45-54, 2019.

**APÊNDICE A – Triagem de repositórios de dados de pesquisa para a composição do universo de pesquisa (re3data).**

Nome do repositório de dados de pesquisa	País (es) a que se vincula	Atua como sistema de autoarquivamento?	Possui e é possível o acesso à políticas e/ou diretrizes para a indexação dos dados de pesquisa?	Qualifica-se para compor o universo da pesquisa?
   <b>BRASIL   QUANTIDADE DE REGISTROS: 21</b>				
<b>WorldClim<sup>35</sup> - Global Climate Data</b>	<i>Brasil (Centro de Referência em Informação Ambiental); outros</i>	Não	Não se aplica	<b>Não</b>
<b>SciELO Data<sup>36</sup></b>	<i>Brasil (Biblioteca Eletrônica Científica Online)</i>	Sim	Sim: “Research data deposit guidelines” <sup>37</sup>	<b>Sim</b>
<b>GLOBE<sup>38</sup></b>	<i>Estados Unidos; Brasil (Instituto Nacional de Pesquisa Espacial, Projeto Global de Terras)</i>	Sim	Sim: Página do repositório - “Case creation” <sup>39</sup> Página do repositório – “Name a Case” <sup>40</sup>	<b>Sim</b>
<b>PPBio Data Repository<sup>41</sup></b>	<i>Brasil (INCT Centro de Estudos Integrados da Biodiversidade Amazônica); outros</i>	Sim	Sim. Recomendações no site do <i>software</i> Morpho para preparação de metadados (especificações do EML - “EML, Cartilha de notação semântica” <sup>42</sup> ) Manual “How to Write Quality Metadata” <sup>43</sup>	<b>Sim</b>
<b>FAPESP COVID-19 Data Sharing/BR<sup>44</sup></b>	<i>Brasil (Fundação de Amparo à Pesquisa de São Paulo; Instituto Fleury; Hospital Israelita Albert Einstein; Hospital Sírio-Libanês; Universidade de São Paulo)</i>	Informação indisponível	Não se aplica	<b>Não</b>
<b>Aleia<sup>45</sup></b>	<i>Brasil (Instituto Brasileiro de Informação em Ciência e Tecnologia)</i>	Sim	Sim: Cartilha de uso do Repositório de Dados de Pesquisa Aleia do Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT) <sup>46</sup>	<b>Sim</b>

<sup>35</sup>Disponível em: <https://worldclim.org/>

<sup>36</sup>Disponível em: <https://data.scielo.org/>

<sup>37</sup>Disponível em: [https://wp.scielo.org/wp-content/uploads/Guia\\_deposito\\_en.pdf](https://wp.scielo.org/wp-content/uploads/Guia_deposito_en.pdf)

<sup>38</sup>Disponível em: <http://globe.umbc.edu/>

<sup>39</sup>Disponível em: <http://globe.umbc.edu/tutorials/case-creation-procedure/#metadata>

<sup>40</sup>Disponível em: <http://globe.umbc.edu/tutorials/how-do-i-name-a-case/>

<sup>41</sup>Disponível em: <https://ppbio.inpa.gov.br/en/repository>

<sup>42</sup>Disponível em: <https://eml.ecoinformatics.org/>

<sup>43</sup>Disponível em: <https://ppbio.inpa.gov.br/sites/default/files/How%20to%20write%20quality%20metadata.pdf>

<sup>44</sup>Disponível em: <https://repositoriodatasharingfapesp.uspdigital.usp.br/>

<sup>45</sup>Disponível em: <https://aleia.ibict.br/>

<sup>46</sup>Disponível em: [https://aleia.ibict.br/assets/pdf/Guia\\_Aleia.pdf](https://aleia.ibict.br/assets/pdf/Guia_Aleia.pdf)

Nome do repositório de dados de pesquisa	País (es) a que se vincula	Atua como sistema de autoarquivamento?	Possui e é possível o acesso à políticas e/ou diretrizes para a indexação dos dados de pesquisa?	Qualifica-se para compor o universo da pesquisa?
Deposita dados <sup>47</sup>	<i>Brasil (Instituto Brasileiro de Informação em Ciência e Tecnologia)</i>	Sim	Sim: Guia de Usuário Deposita Dados <sup>48</sup> Dataverse Project - User Guide <sup>49</sup>	Sim
Base de Dados Científicos da Universidade Federal do Paraná <sup>50</sup>	<i>Brasil (Universidade Federal do Paraná)</i>	Sim	Não	Não
Rede IBICT Cariniana Dataverse <sup>51</sup>	<i>Brasil (Instituto Brasileiro de Informação em Ciência e Tecnologia, Rede Cariniana; Ministério da Ciência, Tecnologia e Inovações)</i>	Sim	Sim: Dataverse Project - User Guide <sup>52</sup>	Sim
Redape <sup>53</sup>	<i>Brasil (Empresa Brasileira de Pesquisa Agropecuária)</i>	Sim	Sim: Guia do usuário – Criação de um conjunto de dados <sup>54</sup> - “Orientações de preenchimento dos metadados <i>citation</i> do Redape”; “Orientações para Catalogação de Datasets Ômicos no Repositório de Dados de Pesquisa da Embrapa <sup>55</sup> ”	Sim
International Ocean Discovery Program <sup>56</sup>	<i>Austrália; Alemanha; Brasil (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior); Índia; Estados Unidos; Japão; República da Coreia; União Europeia; China</i>	Sim	Não	Não
Repositório de Dados de Pesquisa Unifesp <sup>57</sup>	<i>Brasil (Universidade Federal de São Paulo)</i>	Sim	Sim: Dataverse Project - User Guide <sup>58</sup>	Sim

<sup>47</sup>Disponível em: <https://depositadados.ibict.br/>

<sup>48</sup>Disponível em: <https://depositadados.ibict.br/dvn/guide/guides.html>

<sup>49</sup>Disponível em: <https://guides.dataverse.org/en/5.11.1/user/>

<sup>50</sup>Disponível em: <https://bdc.c3sl.ufpr.br/>

<sup>51</sup>Disponível em: <https://repositoriopesquisas.ibict.br/>

<sup>52</sup>Disponível em: <https://guides.dataverse.org/en/5.11.1/user/>

<sup>53</sup>Disponível em: <https://www.redape.dados.embrapa.br/>


<sup>54</sup>Disponível em: <https://www.redape.dados.embrapa.br/guia-do-usuario.xhtml;jsessionid=8d56b967d04414f4bfc98bd413a4#criacao-de-conjunto-de-dados>

<sup>55</sup>Disponível em: <https://www.infoteca.cnptia.embrapa.br/infoteca/bitstream/doc/1138521/1/Doc179-2021.pdf>

<sup>56</sup>Disponível em: <https://www.iodp.org/>

<sup>57</sup>Disponível em: <https://repositoriodedados.unifesp.br/>

<sup>58</sup>Disponível em: <https://guides.dataverse.org/en/5.13/user/>

Nome do repositório de dados de pesquisa	País (es) a que se vincula	Atua como sistema de autoarquivamento?	Possui e é possível o acesso à políticas e/ou diretrizes para a indexação dos dados de pesquisa?	Qualifica-se para compor o universo da pesquisa?
Repositório de Dados de Pesquisa CEDAP - dados de pesquisa <sup>59</sup>	Brasil (Universidade Federal do Rio Grande do Sul)	Informação indisponível	Não se aplica	Não
Repositório Institucional da UNESP <sup>60</sup>	Brasil (Universidade Estadual Paulista)	Sim	Não	Não
Arca Dados <sup>61</sup>	Brasil (Fundação Oswaldo Cruz); outros	Sim	Sim: Dataverse Project - User Guide <sup>62</sup>	Sim
FishSound <sup>63</sup>	Brasil (Universidade de São Paulo); outros	Informação indisponível	Não se aplica	Não
Open Research Data @PUC-Rio <sup>64</sup>	Brasil (Pontifícia Universidade Católica do Rio de Janeiro)	Sim	Não	Não
Dados Abertos De Pesquisas (Repositório de Dados de Pesquisas do Instituto Federal Goiano – Campus Urutai) <sup>65</sup>	Brasil (Instituto Federal de Educação Ciência e Tecnologia Goiano); outros	Sim	Sim. Dataverse Project - User Guide <sup>66</sup>	Sim
REDU - Repositório de Dados de Pesquisa Institucional da Unicamp <sup>67</sup>	Brasil (Universidade de Campinas)	Sim	Sim. Dataverse Project - User Guide <sup>68</sup>	Sim
Maenduar <sup>69</sup>	Brasil (Instituto Brasileiro de Informação em Ciência e Tecnologia; Instituto Brasileiro de Informação em Ciência e Tecnologia, Laboratório em Rede de Humanidades Digitais)	Sim	Sim: Página para <i>upload</i> após <i>login</i>	Sim
Banco de dados de exploração e produção	Brasil (Agência Nacional do Petróleo, Gás Natural e Biocombustíveis; Serviço Geológico do Brasil)	Informação indisponível	Não	Não Indisponível para acesso
 <b>ESPAÑA   QUANTIDADE DE REGISTROS: 55</b>				
UPF Digital Repository <sup>70</sup>	Espanha (Pompeu Fabra University Barcelona)	Sim	Não.	Não

<sup>59</sup>Disponível em: <https://cedap.ufrgs.br/jspui/handle/2050011959/90>

<sup>60</sup>Disponível em: <https://repositorio.unesp.br/home>

<sup>61</sup>Disponível em: <https://arcadados.fiocruz.br/>

<sup>62</sup>Disponível em: <https://guides.dataverse.org/en/5.11.1/user/>

<sup>63</sup>Disponível em: <https://fishsounds.net/index.js>

<sup>64</sup>Disponível em: <https://www.maxwell.vrac.puc-rio.br/projetosEspeciais/ResearchData/index.php?b=1>

<sup>65</sup>Disponível em: <https://dataverse.harvard.edu/dataverse/pesquisa-urt/>

<sup>66</sup>Disponível em: <https://guides.dataverse.org/en/5.11.1/user/>

<sup>67</sup>Disponível em: <https://redu.unicamp.br/>

<sup>68</sup>Disponível em: <https://guides.dataverse.org/en/5.11.1/user/>

<sup>69</sup>Disponível em: <https://zenodo.org/communities/larhud/records?q=&l=list&p=1&s=10>

<sup>70</sup>Disponível em: <https://repositori.upf.edu/>

Nome do repositório de dados de pesquisa	País (es) a que se vincula	Atua como sistema de autoarquivamento?	Possui e é possível o acesso à políticas e/ou diretrizes para a indexação dos dados de pesquisa?	Qualifica-se para compor o universo da pesquisa?
World Values Survey <sup>71</sup>	Espanha (ASEP/JDS); Suécia	Não	Não se aplica	Não
Digibug - Repositorio Institucional de la Universidad de Granada <sup>72</sup>	Espanha (Universidad de Granada)	Sim	Não	Não
International Service of Geomagnetic Indices <sup>73</sup>	França; Dinamarca; Espanha (Ebro Observatory); Alemanha; França; Federação Russa; Japão; outros	Não	Não se aplica	Não
Spanish CLARIN K-Centre <sup>74</sup>	União Europeia; Espanha (vários)	Não	Não se aplica	Não
ioChem-BD <sup>75</sup>	Espanha (Barcelona Supercomputing Center; Institute of Chemical Research of Catalonia; Universitat Rovira i Virgili)	Sim	Sim: Dataverse Project - User Guide <sup>76</sup>	Sim
UCREA - Repositorio Abierto de la Universidad de Cantabria <sup>77</sup>	Espanha (Universidad de Cantabria; Biblioteca Universidad de Cantabria)	Sim	Sim: Guia de autoarchivo <sup>78</sup>	Não Armazena também outros tipos de trabalhos científicos, como Teses e Dissertações
idUS <sup>79</sup> - Deposito de Investigación Universidad de Sevilla	Espanha (Universidad de Sevilla)	Sim	Não	Não
citaREA <sup>80</sup>	Espanha (Centro de Investigación y Tecnología Agroalimentaria de Aragón; Centro de Investigación y Tecnología Agroalimentaria de Aragón, Biblioteca)	Sim	Sim: DSPACE HELP <sup>81</sup>	Não Armazena também outros tipos de trabalhos científicos, como Teses e Dissertações
Repositori de la Universitat Jaume I <sup>82</sup>	Espanha (Universitat Jaume I)	Sim	Não	Não

<sup>71</sup>Disponível em: <https://www.worldvaluessurvey.org/wvs.jsp>

<sup>72</sup>Disponível em: <https://digibug.ugr.es/>

<sup>73</sup>Disponível em: <https://isgi.unistra.fr/>

<sup>74</sup>Disponível em: <http://www.clarin-es-lab.org/index-en.html>

<sup>75</sup>Disponível em: <https://www.iochem-bd.org/>

<sup>76</sup>Disponível em: <https://guides.dataverse.org/en/5.11.1/user/>

<sup>77</sup>Disponível em: <https://repositorio.unican.es/xmlui/>

<sup>78</sup>Disponível em: [https://repositorio.unican.es/xmlui/themes/unican/lib/guia\\_autoarchivo\\_ucrea.pdf](https://repositorio.unican.es/xmlui/themes/unican/lib/guia_autoarchivo_ucrea.pdf)

<sup>79</sup>Disponível em: <https://idus.us.es/>

<sup>80</sup>Disponível em: <https://citatea.cita-aragon.es/citatea/>

<sup>81</sup>Disponível em: <https://citatea.cita-aragon.es/citatea/help/index.html>

<sup>82</sup>Disponível em: <https://repositori.uji.es/xmlui/>

Nome do repositório de dados de pesquisa	País (es) a que se vincula	Atua como sistema de autoarquivamento?	Possui e é possível o acesso à políticas e/ou diretrizes para a indexação dos dados de pesquisa?	Qualifica-se para compor o universo da pesquisa?
Tools for Second Language Speech Research and Teaching <sup>83</sup>	Reino Unido; Japão; <i>Espanha (Universidad de Barcelona)</i>	Não	Não se aplica	Não
Repositorio Abierto - Repositorio Abierto de la Universidad Internacional de Andalucía <sup>84</sup>	<i>Espanha (Universidad Internacional de Andalucía)</i>	Sim	Informação indisponível	Não
The Perovskite Database Project <sup>85</sup>	<i>Espanha (GRECO); Alemanha; Israel</i>	Sim	Não	Não
Dados MIPAS/Envisat gerados pela IMK/IAA <sup>86</sup>	<i>Espanha (Instituto de Astrofísica da Andaluzia); Alemanha</i>	Informação indisponível	Não se aplica	Não
Dipòsit Digital de la Universitat de Barcelona Dades <sup>87</sup>	<i>Espanha (Universitat de Barcelona)</i>	Sim	Informação indisponível	Não
Centro de Investigaciones Sociológicas Data Bank y Estudios <sup>88</sup>	<i>Espanha (Centro de Investigaciones Sociológicas)</i>	Informação indisponível	Não se aplica	Não
O2 Repositori UOC - Dades <sup>89</sup>	<i>Espanha (Universitat Oberta de Catalunya)</i>	Sim	Informação indisponível	Não
MultiDark Database <sup>90</sup>	<i>Alemanha; Espanha (Multimessenger Approach for Dark Matter Detection); Bélgica</i>	Não	Não se aplica	Não
UvaDOC - Fonda Antigo <sup>91</sup> ; Repositório Documental da Universidade de Valladolid - Fonda Antigo	<i>Espanha (Universidad de Valladolid)</i>	Informação indisponível	Não se aplica	Não
Biblioteca de Datos de CEACS <sup>92</sup>	<i>Espanha (Fundación Juan March; Instituto Carlos III-Juan March, Centro de investigación y posgrado en Ciencias Sociales)</i>	Não	Não se aplica	Não

<sup>83</sup>Disponível em: <http://sla-speech-tools.com/>

<sup>84</sup>Disponível em: <https://dspace.unia.es/>

<sup>85</sup>Disponível em: <https://www.perovskitedatabase.com/>

<sup>86</sup>Disponível em: <https://www.imk-asf.kit.edu/english/308.php>

<sup>87</sup>Disponível em: <https://diposit.ub.edu/dspace/handle/2445/56364>

<sup>88</sup>Disponível em: <https://www.cis.es/estudios/catalogo-estudios>

<sup>89</sup>Disponível em: <https://openaccess.uoc.edu/>

<sup>90</sup>Disponível em: <https://www.multidark.org/>

<sup>91</sup>Disponível em: <https://uvadoc.uva.es/handle/10324/150>

<sup>92</sup>Disponível em: <http://www.march.es/ceacs/biblioteca/dataib/>

Nome do repositório de dados de pesquisa	País (es) a que se vincula	Atua como sistema de autoarquivamento?	Possui e é possível o acesso à políticas e/ou diretrizes para a indexação dos dados de pesquisa?	Qualifica-se para compor o universo da pesquisa?
Antimicrobial Combination Networks <sup>93</sup>	Espanha (Sing Group); Portugal	Não	Não se aplica	Não
FactGrid <sup>94</sup>	Espanha (Fundación Ignacio Larramendi; Universitat Pompeu Fabra; Universitat de Barcelona); Alemanha; França; Estados Unidos	Sim	Não	Não
SIOR <sup>95</sup> - Social Impact of Science	Espanha (Universitat de Barcelona)	Informação indisponível	Informação indisponível	Não Indisponível para acesso
Protected Planet <sup>96</sup>	Espanha (vizzuality); outros	Não	Não se aplica	Não
Analysis of the Interstellar Medium of Isolated Galaxies - AMIGA	Espanha (Consejo Superior de Investigaciones Científicas; Fundación Centro de Supercomputación de Castilla Y Leon; Institute of Astrophysics of Andalusia)	Não	Não se aplica	Não
DIGITAL.CSIC <sup>97</sup>	Espanha (Gobierno de España, Ministerio de Economía y Competitividad, Secretaría de Estado de Investigación, Desarrollo e Innovación; Spanish National Research Council)	Sim	Não	Não
Migrant Integration Policy Index <sup>98</sup>	Espanha (Barcelona Centre for International Affairs); outros	Não	Não se aplica	Não
RETOPEA <sup>99</sup>	Espanha (Euro-Arab Foundation for Higher Studies; Universidad de Granada); Bélgica; Alemanha; Macedônia; Reino Unido; Finlândia; Estônia; Polônia	Não	Não se aplica	Não
Repositorio de Activos Digitales del Instituto Andaluz de Patrimonio Histórico	Espanha (Andalusian Institute of Historical Heritage)	Sim	Não	Não
Repositorio Institucional de la Universidad de Burgos <sup>100</sup>	Espanha (Universidad de Burgos; Biblioteca Universitaria Universidad de Burgos); União Europeia	Sim	Não	Não
Repositorio Institucional del Instituto Madrileño de Estudios Avanzados en Nanociencia <sup>101</sup>	Espanha (Madrid Institute for Advanced Studies; Instituto Madrileño de Estudios Avanzados en Nanociencia)	Informação indisponível	Informação indisponível	Não

<sup>93</sup>Disponível em: <https://www.sing-group.org/antimicrobialCombination/networks/view>

<sup>94</sup>Disponível em: [https://database.factgrid.de/wiki/Main\\_Page](https://database.factgrid.de/wiki/Main_Page)

<sup>95</sup>Disponível em: <https://www.ub.edu/sior/>

<sup>96</sup>Disponível em: <https://www.protectedplanet.net/en>

<sup>97</sup>Disponível em: <https://digital.csic.es/?locale=en>

<sup>98</sup>Disponível em: <http://www.mipex.eu/>

<sup>99</sup>Disponível em: <https://www.retopea.eu/>

<sup>100</sup>Disponível em: <https://riubu.ubu.es/>

<sup>101</sup>Disponível em: <https://repositorio.imdeananociencia.org/>

Nome do repositório de dados de pesquisa	País (es) a que se vincula	Atua como sistema de autoarquivamento?	Possui e é possível o acesso à políticas e/ou diretrizes para a indexação dos dados de pesquisa?	Qualifica-se para compor o universo da pesquisa?
<b>RIUMA<sup>102</sup> - Repositorio Institucional de la Universidad de Málaga</b>	<i>Espanha (Universidad de Málaga)</i>	Sim	Sim: Manual de uso > Cómo depositar un archivo > Descripción del ítem <sup>103</sup>	<b>Não</b> Armazena também outros tipos de trabalhos científicos, como Teses e Dissertações
<b>Italian Centre for Astronomical Archive<sup>104</sup></b>	<i>Espanha (Fundación Galileo Galilei - INAF, Telescopio Nazionale Galileo, Fundación Canaria); Itália</i>	Não	Não se aplica	<b>Não</b>
<b>Arias Montano<sup>105</sup> - Repositorio Institucional da Universidade de Huelva</b>	<i>Espanha (Universidad de Huelva)</i>	Sim	Sim: Página do repositório - Depositar un documento > Empezar un depósito : proceso de envío <sup>106</sup> ; Ayuda DSpace <sup>107</sup>	<b>Não</b> Armazena também outros tipos de trabalhos científicos, como Teses e Dissertações
<b>UPCommons<sup>108</sup> - Research Data</b>	<i>Espanha (Universitat Politècnica de Catalunya)</i>	Informação indisponível	Não se aplica	<b>Não</b>
<b>CORA. Reposeitori de Dades de Recerca<sup>109</sup></b>	<i>Espanha (vários)</i>	Sim	Sim: Página do repositório - User Guide > Crear y depositar un dataset <sup>110</sup> ; GUIA - Guía práctica para la publicación de datos tabulares en archivos - CSV <sup>111</sup>	<b>Sim</b>
<b>MindBigData<sup>112</sup></b>	<i>Espanha (MindBigData.com)</i>	Não	Não se aplica	<b>Não</b>
<b>Gran Telescopio CANARIAS Public Archive<sup>113</sup></b>	<i>Espanha (Centro de Astrobiología, Instituto Nacional de Técnica Aeroespacial; Gran Telescopio Canarias S.A.; Instituto de Astrofísica de Canarias); União Europeia; México; Estados Unidos</i>	Informação indisponível	Informação indisponível	<b>Não</b>

<sup>102</sup>Disponível em: <https://riuma.uma.es/>

<sup>103</sup>Disponível em: <https://biblioguias.uma.es/RIUMA/Descripcion>

<sup>104</sup>Disponível em: <https://www.ia2.inaf.it/>

<sup>105</sup>Disponível em: <https://rabida.uhu.es/dspace/>

<sup>106</sup>Disponível em: <https://guiasbuh.uhu.es/c.php?g=463635&p=4016656#s-lg-box-wrapper-14713669>

<sup>107</sup>Disponível em: <https://rabida.uhu.es/dspace/ayuda.html>

<sup>108</sup>Disponível em: <https://upcommons.upc.edu/dades>

<sup>109</sup>Disponível em: <https://www.re3data.org/repository/https://dataverse.csuc.cat/>

<sup>110</sup>Disponível em: <https://confluence.csuc.cat/display/RDM/Crear+i+dipositar+un+dataset>

<sup>111</sup>Disponível em: [https://urv.libguides.com/ld.php?content\\_id=34503012](https://urv.libguides.com/ld.php?content_id=34503012)

<sup>112</sup>Disponível em: <http://www.mindbigdata.com/>

<sup>113</sup>Disponível em: <https://gtc.sdc.cab.inta-csic.es/gtc/>

Nome do repositório de dados de pesquisa	País (es) a que se vincula	Atua como sistema de autoarquivamento?	Possui e é possível o acesso à políticas e/ou diretrizes para a indexação dos dados de pesquisa?	Qualifica-se para compor o universo da pesquisa?
RiuNet	Espanha ( <i>Universitat Politècnica de València; Universitat Politècnica de València, Servicio de Biblioteca y Documentación Científica</i> )	Sim	Sim: Página do repositório - Deposit <sup>114</sup>	Sim
ICTS SOCIB Data Repository <sup>115</sup>	Espanha ( <i>Govern de les Illes Balears; Ministerio de Ciencia e Innovación; Consejo Superior de Investigaciones Científicas</i> )	Não	Não se aplica	Não
Dipòsit Digital de Documents de la UAB <sup>116</sup>	Espanha ( <i>Universitat Autònoma de Barcelona</i> )	Sim	Não	Não
DisGeNET <sup>117</sup>	Espanha ( <i>Research Programme on Biomedical Informatics, Integrative Biomedical Informatics</i> )	Informação indisponível	Informação indisponível	Não
UTM-CSIC Data Centre <sup>118</sup>	Espanha ( <i>Unidade de Tecnologia Marina; Conselho Nacional de Pesquisa Espanhol</i> )	Não	Não se aplica	Não
AHEAD - European Archive of Historical Earthquake Data <sup>119</sup>	França; União Europeia; Espanha ( <i>Institut Cartogràfic i Geològic de Catalunya</i> ); França; Grécia; Portugal; Itália; Bélgica; Suíça; Noruega; Finlândia	Não	Não se aplica	Não
Constrained Local UniversE Simulations <sup>120</sup>	Espanha ( <i>Centro Nacional de Supercomputacion; Universidad Autonómica de Madrid, Grupo de Astrofísica</i> ); Alemanha; Israel; México	Não	Não se aplica	Não
The European Genome-phenome Archive <sup>121</sup>	Espanha ( <i>Centre for Genomic Regulation</i> ); União Europeia	Sim	Não	Não
eBiltegia <sup>122</sup>	Espanha ( <i>Universidad de Mondragón</i> )	Sim	Informação indisponível	Não
APID <sup>123</sup> - Agile Protein Interactomes DataServer	Espanha ( <i>Centro de Investigacion del Cancer, Instituto Biología Molecular y Celular del Cancer; Consejo Superior de Investigaciones Científicas; Instituto de Salud Carlos III; Universidad de Salamanca</i> )	Não	Não se aplica	Não
UdG Digital Repository <sup>124</sup>	Espanha ( <i>Universitat de Girona</i> )	Sim	Informação indisponível	Não

<sup>114</sup>Disponível em: <https://riunet.upv.es/help/depositar>

<sup>115</sup>Disponível em: <https://www.socib.es/data/>

<sup>116</sup>Disponível em: <https://ddd.uab.cat/>

<sup>117</sup>Disponível em: <https://www.disgenet.org/>

<sup>118</sup>Disponível em: <http://data.utm.csic.es/>

<sup>119</sup>Disponível em: <https://www.emidius.eu/AHEAD/>


<sup>120</sup>Disponível em: <https://www.clues-project.org/cms/>

<sup>121</sup>Disponível em: <https://ega-archive.org/>

<sup>122</sup>Disponível em: <https://ebiltegia.mondragon.edu/>

<sup>123</sup>Disponível em: <http://cicblade.dep.usal.es:8080/APID/init.action>

<sup>124</sup>Disponível em: <https://dugi-doc.udg.edu/>

Nome do repositório de dados de pesquisa	País (es) a que se vincula	Atua como sistema de autoarquivamento?	Possui e é possível o acesso à políticas e/ou diretrizes para a indexação dos dados de pesquisa?	Qualifica-se para compor o universo da pesquisa?
RIULL <sup>125</sup> - Repositorio Institucional de la Universidad de La Laguna	Espanha (Universidad de La Laguna)	Informação indisponível	Informação indisponível	Não
Phenol-Explorer <sup>126</sup>	Espanha (Universitat de Barcelona); outros	Sim	Não	Não
CosmoHub <sup>127</sup>	Espanha (Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas; Institut de Física d'Altes Energies; Institute of Space Sciences; Port d'Informació Científica)	Sim	Informação indisponível	Não
CEU Institutional Repository <sup>128</sup>	Espanha (Fundación Universitaria San Pablo CEU)	Sim	Sim: Ayuda DSpace <sup>129</sup>	Não Armazena também outros tipos de trabalhos científicos, como Teses e Dissertações
Repositori Institucional Universitat Rovira i Virgili <sup>130</sup>	Espanha (Universitat Rovira i Virgili)	Sim	Sim: Página do repositório - User Guide > Crear y depositar un dataset <sup>131</sup> (indica valores controlados); Guia - Guía práctica para la publicación de datos tabulares en archivos - CSV <sup>132</sup>	Não Armazena também outros tipos de trabalhos científicos, como Teses e Dissertações
e-cienciaDatos <sup>133</sup>	Espanha (vários); outros	Sim	Informação indisponível	Não
   PORTUGAL   QUANTIDADE DE REGISTROS: 15				
DUnAs <sup>134</sup> - Repositório de dados de investigação da Universidade de Aveiro	Portugal (Universidade de Aveiro)	Sim	Sim: Dataverse Project - User Guide; <sup>135</sup> Metadados: bloco "citation metadata" <sup>136</sup>	Sim

<sup>125</sup>Disponível em: <https://riull.ull.es/xmlui/>

<sup>126</sup>Disponível em: <http://phenol-explorer.eu/>

<sup>127</sup>Disponível em: <https://cosmohub.pic.es/>

<sup>128</sup>Disponível em: <https://repositorioinstitucional.ceu.es/>

<sup>129</sup>Disponível em: <https://repositorioinstitucional.ceu.es/help/index.html>

<sup>130</sup>Disponível em: <https://www.crai.urv.cat/en/information-resources/repository/>

<sup>131</sup>Disponível em: <https://confluence.csuc.cat/display/RDM/Crear+i+dipositar+un+dataset>

<sup>132</sup>Disponível em: [https://urv.libguides.com/ld.php?content\\_id=34503012](https://urv.libguides.com/ld.php?content_id=34503012)

<sup>133</sup>Disponível em: <https://edatos.consorciomadrono.es/>

<sup>134</sup>Disponível em: <https://dunas.ua.pt/>

<sup>135</sup>Disponível em: <https://guides.dataverse.org/en/5.11/user/>

<sup>136</sup>Disponível em: <https://uapt.libguides.com/dunas/criar-dataset>

Nome do repositório de dados de pesquisa	País (es) a que se vincula	Atua como sistema de autoarquivamento?	Possui e é possível o acesso à políticas e/ou diretrizes para a indexação dos dados de pesquisa?	Qualifica-se para compor o universo da pesquisa?
<b>APIS<sup>137</sup> - Arquivo Português de Informação Social</b>	<i>Portugal (Fundação para a Ciência e a Tecnologia; Universidade de Lisboa, Instituto de Ciências Sociais)</i>	Sim	Sim: Página do repositório "Preparar os metadados" <sup>138</sup> ; Ajuda do Repositório <sup>139</sup>	Sim
<b>Biofilms Structural Database<sup>140</sup></b>	<i>Portugal (Universidade do Porto, Faculdade de Medicina, Departamento de Biomedicina)</i>	Sim	Informação indisponível	Não
<b>Antimicrobial Combination Networks<sup>141</sup></b>	<i>Espanha; Portugal (Universidade do Minho, Centro de Engenharia Biológica)</i>	Não	Não se aplica	Não
<b>PORTULAN CLARIN repository<sup>142</sup></b>	<i>União Europeia; Portugal (Universidade de Lisboa, Faculdade de Ciências da Universidade de Lisboa, Departamento de Informática)</i>	Sim	Não	Não
<b>Arquivo.pt<sup>143</sup></b>	<i>Portugal (Fundação para a Ciência e a Tecnologia)</i>	Não	Não se aplica	Não
<b>Repositório de Dados Científicos<sup>144</sup></b>	<i>União Europeia; Portugal (Fundação para a Ciência e a Tecnologia; Project Partner; Repositórios Científicos de Acesso Aberto de Portugal; República Portuguesa)</i>	Sim	Sim: Página do repositório "Preparar os metadados" <sup>145</sup> ; Ajuda do Repositório <sup>146</sup>	Sim
<b>LegionellaDB<sup>147</sup></b>	<i>Portugal (Universidade do Porto, Faculdade de Medicina, Departamento de Biomedicina)</i>	Não	Não se aplica	Não
<b>KiMoSys - Kinetic models of biological systems<sup>148</sup></b>	<i>União Europeia; Portugal (Fundação para a Ciência e a Tecnologia; Instituto de Engenharia de Sistemas e Computadores, Investigação e Desenvolvimento em Lisboa)</i>	Informação indisponível.	Informação indisponível	Não Indisponível para acesso
<b>DadosIPB<sup>149</sup></b>	<i>Portugal (Instituto Politécnico de Bragança)</i>	Sim	Sim: Dataverse Project - User Guide <sup>150</sup>	Sim

<sup>137</sup>Disponível em: <https://dados.rcaap.pt/handle/10400.20/1>

<sup>138</sup>Disponível em: <http://www.apis.ics.ulisboa.pt/metadados/>

<sup>139</sup>Disponível em: [https://dados.rcaap.pt/help\\_pt/index.html](https://dados.rcaap.pt/help_pt/index.html)

<sup>140</sup>Disponível em: <https://biofilms.biosim.pt/>

<sup>141</sup>Disponível em: <https://www.sing-group.org/antimicrobialCombination/networks/view>

<sup>142</sup>Disponível em: <https://portulanclarin.net/>

<sup>143</sup>Disponível em: <https://arquivo.pt/>

<sup>144</sup>Disponível em: <https://dados.rcaap.pt/?locale=en>

<sup>145</sup>Disponível em: <http://www.apis.ics.ulisboa.pt/metadados/>

<sup>146</sup>Disponível em: [https://dados.rcaap.pt/help\\_pt/index.html](https://dados.rcaap.pt/help_pt/index.html)

<sup>147</sup>Disponível em: <https://legionelladb.biosim.pt/>

<sup>148</sup>Disponível em: <https://kimosys.org/>

<sup>149</sup>Disponível em: <https://dados.ipb.pt/>

<sup>150</sup>Disponível em: <https://guides.dataverse.org/en/5.11.1/user/>

Nome do repositório de dados de pesquisa	País (es) a que se vincula	Atua como sistema de autoarquivamento?	Possui e é possível o acesso à políticas e/ou diretrizes para a indexação dos dados de pesquisa?	Qualifica-se para compor o universo da pesquisa?
AHEAD - European Archive of Historical Earthquake Data <sup>151</sup>	União Europeia; Espanha; França; Grécia; <i>Portugal (Instituto Português do Mar e da Atmosfera)</i> ; Itália; Bélgica; Suíça; Noruega; Finlândia	Não	Não se aplica	Não
INTEGRALL <sup>152</sup>	França; <i>Portugal (Universidade de Aveiro, Centro de estudos do ambiente e do mar; Universidade de Aveiro, Departamento de Biologia)</i>	Não	Não se aplica	Não
Perdigao Field Experiment <sup>153</sup>	União Europeia; <i>Portugal (Universidade do Porto, Faculdade de Engenharia)</i>	Sim	Sim: Página do repositório - Diretrizes para Documentação de Conjunto de Dados <sup>154</sup> ; Instruções para envio de dados da Perdigão <sup>155</sup>	Sim
DataRepositoriUM <sup>156</sup> - Repositório de Dados da Universidade do Minho	<i>Portugal (Universidade de Minho)</i>	Sim	Sim: Dataverse Project - User Guide <sup>157</sup>	Sim
INESC TEC Research Data Repository <sup>158</sup>	<i>Portugal (INESC TEC)</i>	Informação indisponível	Informação indisponível	Não

Fonte: elaboração própria (2024)

<sup>151</sup>Disponível em: <https://www.emidius.eu/AHEAD/>

<sup>152</sup>Disponível em: <http://integrall.bio.ua.pt/>

<sup>153</sup>Disponível em: <https://perdigao.fe.up.pt/>

<sup>154</sup>Disponível em: <https://www.eol.ucar.edu/content/dataset-documentation-requirements>

<sup>155</sup>Disponível em: <https://www.eol.ucar.edu/content/perdigao-data-submission-instructions>

<sup>156</sup>Disponível em: <https://datarepositorium.uminho.pt/>

<sup>157</sup>Disponível em: <https://guides.dataverse.org/en/5.11.1/user/>

<sup>158</sup>Disponível em: <https://rdm.inesctec.pt/>