



UNIVERSIDADE
ESTADUAL DE LONDRINA

FELIPE CALDONAZZO DE ALMEIDA PEREIRA

**ORGANIZAÇÃO E REPRESENTAÇÃO DO CONJUNTO
DOCUMENTAL EM REPOSITÓRIOS DIGITAIS INTEGRADOS
DA AMÉRICA DO SUL**

Londrina
2022

FELIPE CALDONAZZO DE ALMEIDA PEREIRA

**ORGANIZAÇÃO E REPRESENTAÇÃO DO CONJUNTO
DOCUMENTAL EM REPOSITÓRIOS DIGITAIS INTEGRADOS
DA AMÉRICA DO SUL**

Dissertação de mestrado apresentada ao Programa de Pós-Graduação em Ciência da Informação da Universidade Estadual de Londrina, como requisito parcial à obtenção do título de mestre em Ciência da Informação. Linha de pesquisa: Organização e Representação da Informação e do Conhecimento.

Orientadora: Prof^a. Dr^a. Brígida Maria Nogueira Cervantes.

Londrina
2022

Ficha de identificação da obra elaborada pelo autor, através do Programa de Geração Automática do Sistema de Bibliotecas da UEL

P436 Pereira, Felipe Caldonazzo de Almeida.
Organização e representação do conjunto documental nos repositórios digitais integrados da América do Sul / Felipe Caldonazzo de Almeida Pereira. - Londrina, 2022.
78 f.

Orientador: Brígida Maria Nogueira Cervantes.
Dissertação (Mestrado em Ciência da Informação) - Universidade Estadual de Londrina, Centro de Educação Comunicação e Artes, Programa de Pós-Graduação em Ciência da Informação, 2022.
Inclui bibliografia.

1. Organização da informação - Tese. 2. Representação da informação - Tese. 3. Comunicação científica - Tese. 4. Metadados - Tese. I. Cervantes, Brígida Maria Nogueira. II. Universidade Estadual de Londrina. Centro de Educação Comunicação e Artes. Programa de Pós-Graduação em Ciência da Informação. III. Título.

CDU 02

FELIPE CALDONAZZO DE ALMEIDA PEREIRA

**ORGANIZAÇÃO E REPRESENTAÇÃO DO CONJUNTO
DOCUMENTAL EM REPOSITÓRIOS DIGITAIS INTEGRADOS
DA AMÉRICA DO SUL**

Dissertação de mestrado apresentada ao Programa de Pós-Graduação em Ciência da Informação da Universidade Estadual de Londrina, como requisito parcial à obtenção do título de mestre em Ciência da Informação. Linha de pesquisa: Organização e Representação da Informação e do Conhecimento.

BANCA EXAMINADORA

Orientadora: Prof^a. Dr^a. Brígida Maria Nogueira
Cervantes
Universidade Estadual de Londrina – UEL

Prof. Dr. Miguel Luiz Contani
Universidade Estadual de Londrina – UEL

Prof^a Dr^a. Paula Regina Ventura Amorim
Gonzalez
Universidade Federal do Espírito Santo - UFES

Londrina 2 de maio de 2022.

AGRADECIMENTOS

Primeiramente, um agradecimento especial à minha mãe, por estar sempre presente e disponível para ajudar no que fosse possível e impossível, principalmente nas minhas tomadas de decisões que afetariam a minha vida. Em segundo, à minha querida orientadora, nesta parceria que vem desde o primeiro ano de minha graduação, na disciplina de Esquemas de Classificação, me convidando a participar de seu grupo de pesquisa, me orientando na iniciação científica e no trabalho de conclusão de curso, culminando na orientação no mestrado, meu eterno agradecimento e admiração.

Um agradecimento especial aos amigos Nicolas e Rose, cada um em sua área e especialidade, pelas conversas e discussões sobre a vida e a pesquisa, além dos projetos de produções futuras que virão.

Aos amigos pelas boas conversas relacionadas às experiências no mestrado, entre outros assuntos e que de cada forma puderam me ajudar: Geneviane, Géssi, Guto e Tai. E um agradecimento especial à Flávia, que nesses últimos meses me emprestou sua orelha e atenção.

Aos participantes do grupo de pesquisa “Vocabulário controlado no portal de periódicos científicos eletrônicos da Universidade Estadual de Londrina: foco na gestão dos conteúdos temáticos dos vocabulários controlados”, as reuniões das sextas-feiras tiveram boas discussões que foram bem aproveitadas.

Aos docentes Programa de Pós-Graduação em Ciência da Informação da UEL (PPGCI-UEL), as disciplinas foram adaptadas para o modo remoto, mas a dedicação e o entusiasmo dos participantes compensaram a distância. Aos servidores da secretaria do CECA e da PROPPG por ajudarem no que fosse possível nesse período de pandemia (que ainda não acabou!). E à CAPES pelo financiamento desta pesquisa.

“Nosso tema é o óbvio. Acho mesmo que os cientistas trabalham com o óbvio. O negócio deles – nosso negócio – é lidar com o óbvio. Aparentemente, Deus é muito treteiro, faz as coisas de forma tão recôndita e disfarçada que se precisa desta categoria de gente – os cientistas – para ir tirando os véus, desvendando, a fim de revelar a obviedade do óbvio. O ruim deste procedimento é que parece um jogo sem fim. De fato, só conseguimos desmascarar uma obviedade para descobrir outras, mais óbvias ainda.” (Darcy Ribeiro)

PEREIRA, F. C. A. **Organização e representação do conjunto documental em repositórios digitais integrados da América do Sul.** 2022. 78 f. Dissertação (Mestrado em Ciência da Informação) – Departamento de Ciência da Informação, Universidade Estadual de Londrina, Londrina, 2022.

RESUMO

Introdução: A Ciência Aberta tem como uma de suas características o conjunto de práticas que permitem o acesso aberto aos procedimentos e resultados de uma pesquisa científica, isso inclui tanto a publicação científica quanto a disponibilização dos dados de pesquisa, fazendo-se necessária a análise de diferentes formas para se organizar e representar a informação científica, entre estas, a do conjunto documental no âmbito do repositório digital integrado. **Objetivo:** este trabalho tem como objetivo geral analisar o uso de recursos instrumentais para organização e representação do conjunto documental em repositório digital integrado. Os objetivos específicos são a identificação das características do repositório digital integrado, a verificação das características do conjunto documental e a seleção de recursos para organização e representação do conjunto documental aplicáveis ao contexto do repositório digital integrado. **Metodologia:** esta pesquisa é de natureza descritiva, utilizando-se de levantamento bibliográfico e documental, com uso da Análise de Conteúdo para a identificação e a seleção dos documentos nos repositórios digitais e para a categorização dos elementos coletados. A coleta de dados para a seleção dos repositórios foi a partir da lista existente nos diretórios OPENDOAR e ROAR, sendo identificados 32 repositórios com as características de repositórios digitais integrados. **Resultado:** tem-se que os 32 repositórios utilizam ou fazem indicação para o depósito de arquivos em formato aberto ou não proprietário, somente um repositório não utiliza recursos e práticas para dados conectados e nove repositórios não relacionam os objetos pertencentes ao conjunto documental. Apesar do uso de vocabulário controlado (assunto e tipologia documental), somente 11 repositórios empregam os recursos de dados conectados para esse fim. **Conclusão:** a filosofia do movimento de acesso aberto possui aceitação nas instituições responsáveis pelos repositórios digitais integrados, porém, percebeu-se a falta de padronização no emprego dos campos de metadados e no uso de recursos e práticas voltadas para a utilização de dados abertos conectados recomendados nas diretrizes dos órgãos responsáveis pela formulação de políticas de informação científica.

Palavras-chave: organização da informação; representação da informação; comunicação científica; publicação ampliada; metadados.

PEREIRA, F. C. A. **Organization and representation of the documental set in digital integrated repositories of South America**. 2022. 78 p. Dissertation (Master in Information Science) – State University of Londrina, Londrina, 2022.

ABSTRACT

Introduction: Open Science has as one its characteristics the set of practices that allow the open access to procedures and results of scientific research, this includes both the scientific publication and the availability of research data, which makes the analysis of different forms of organizing and representing scientific information necessary, among these, the documental set in the framework of digital integrated repository. Objective: this work has the general objective of analysing the use of instrumental resources for the organization and representation of documental sets in integrated digital repository. Specific objectives are the identification of characteristics of the integrated digital repository; the verification of the documental set; and the investigation of the use of metadata schemes for the organization and representation of documental sets. Methodology: this is a descriptive research, using both bibliographical and documental research, with Content Analysis for the identification and selection of documents in digital repositories and for the categorization of collected elements. The data collection for the selection of repositories departed from the pre-existing list on OPENDOAR and ROAR's directories, of which 32 repositories with the characteristics of digital integrated repositories were identified. Results: there are 32 repositories that utilize or make indication for the depository of archives in open format or non-proprietary, of which one does not utilize resources and practices for connected data, and nine repositories do not relate the pertaining objects to the documental set. Despite the use of controlled vocabulary (subject and documental tipology), only 11 repositories employed the resources of linked data to that purpose. Conclusion: the philosophy of the open access movement has great acceptance on institutions that are responsible for digital integrated repositories; however, it is perceived the lack of pattern on the employment of fields of metadata and on the use of resources and practices aimed at the utilization of linked open data recommended on the guidelines of responsible agencies for the formulation of scientific information politics.

Key words: information organization; information representation; scientific communication; enhanced publication; medatada.

LISTA DE FIGURAS

Figura 1 -	O conjunto documental no ciclo de vida da pesquisa.....	40
Figura 2 -	Página inicial do Repositório Digital ANII	47
Figura 3 -	Busca avançada Repositório FCEyS-UNMDP	51
Figura 4 -	Vocabulário da Biblioteca Central da FFyL	57
Figura 5 -	Termos específicos de Bibliotecologia	58
Figura 6 -	Nota bibliográfica e metadados.....	59
Figura 7 -	Classificação FORD	60
Figura 8 -	Ciencias de la computación e información	61
Figura 9 -	Tesouro AGROVOC.....	62
Figura 10 -	AGROVOC metadados	62
Figura 11 -	Tipologia para recursos COAR	64

LISTA DE QUADROS

Quadro 1 – Sistematização metodológica da pesquisa	17
Quadro 2 – Funções e descrição das funções de metadados	24
Quadro 3 – Aspectos da interoperabilidade	30
Quadro 4 – Repositórios da América do Sul	45
Quadro 5 – Características dos repositórios da América do Sul	47
Quadro 6 – Dados abertos conectados e outros recursos	52
Quadro 7 – PURL da tipologia documental / tipo de recurso	63
Quadro 8 – Tipologias info:ar-repo, info:eu-repo e info:pe-repo	65
Quadro 9 – Campos de metadados e o conjunto documental	68
Quadro 10 – Campos que utilizam dados conectados	69

LISTA DE ABREVIATURAS E SIGLAS

AACR2	Anglo-American Cataloguing Rules 2.ed.
CNPq	Conselho Nacional de Desenvolvimento Científico e Tecnológico
COAR	Confederation of Open Access Repositories
DC	Dublin Core
DRIVER	Digital Repositories Infrastructure Vision for European Research
DOI	Digital Object Identifier
FAIR	Findable, Accessible, Interoperable and Reusable Data
HTTP	Hypertext Transfer Protocol
IFLA	International Federation of Library Associations and Institutions
LOM	Learning Object Metadata
LRM	Library Reference Model
OAI	Open Archives Initiative
OAI-PMH	Open Archives Initiative Protocol for Metadata Harvesting
OAIRE	Open Access Infrastructure for Research in Europe
ODF	Open Document Format
ORCID	Open Researcher and Contributor
PDF	Portable Document Format
PURL	Persistent Uniform Resource Locator
RDF	Resource Description Framework
RDFS	Resource Description Framework Schema
REDCOL	Rede Colombiana de Informação Científica
RENATI	Registro Nacional de Trabajos de Investigación
SNRD	Sistema Nacional de Repositórios Digitais
SOC	Sistema de Organização do Conhecimento
URI	Uniform Resource Identifier
URN	Uniform Resource Name
VIAF	Virtual International Authority File
XML	Extensible Markup Language
W3C	World Wide Web Consortium

SUMÁRIO

1	INTRODUÇÃO	12
1.1	PROBLEMATIZAÇÃO	14
1.2	JUSTIFICATIVA.....	15
1.3	OBJETIVOS	15
1.3.1	Objetivo Geral.....	17
1.3.2	Objetivos Específicos	17
1.4	SISTEMATIZAÇÃO METODOLÓGICA DA PESQUISA	17
1.5	ESTRUTURA DA PESQUISA	18
2	ORGANIZAÇÃO E REPRESENTAÇÃO DA INFORMAÇÃO	19
2.1	ORGANIZAÇÃO E REPRESENTAÇÃO DA INFORMAÇÃO: CONCEITUAÇÃO	19
2.2	RECURSOS PARA REPRESENTAÇÃO DA INFORMAÇÃO EM AMBIENTE DIGITAL ..	21
2.2.1	Dados Abertos Conectados	21
2.2.2	Esquemas de Metadados	24
2.2.3	Identificadores Persistentes de Autoria	28
2.2.4	Interoperabilidade	29
3	COMUNICAÇÃO CIENTÍFICA: DOS PERIÓDICOS CIENTÍFICOS AOS DADOS DE PESQUISA	32
3.1	COMUNICAÇÃO CIENTÍFICA: ORIGENS E DISCUSSÕES	32
3.2	CIÊNCIA ABERTA	33
3.3	DADOS DE PESQUISA	34
3.4	REPOSITÓRIOS DE DADOS DE PESQUISA	35
3.5	REPOSITÓRIO DIGITAL INTEGRADO	36
3.6	CONJUNTO DOCUMENTAL	37
4	PROCEDIMENTOS METODOLÓGICOS	41
5	ANÁLISE DOS DADOS E RESULTADOS	44
5.1	REPOSITÓRIOS DIGITAIS INTEGRADOS DA AMÉRICA DO SUL.....	45
5.2	AS PRÁTICAS E OS RECURSOS PARA REPRESENTAÇÃO E ORGANIZAÇÃO DO CONJUNTO DOCUMENTAL	51

5.2.1	Dados Abertos Conectados e Identificados Persistentes	54
5.2.2	Vocabulários Controlados.....	56
5.2.3	Conjunto Documental	65
6	CONSIDERAÇÕES FINAIS	71
	REFERÊNCIAS	73

1 INTRODUÇÃO

O movimento *Open Science* (Ciência Aberta) é uma cultura ou filosofia que tem a intenção de tornar acessível a um maior número de pessoas (pesquisadores ou não) o fazer científico. Nesse sentido, é importante levar em consideração algumas de suas características, como o conjunto de práticas que permitem o acesso aberto aos procedimentos, metodologias, discussão, validação e resultado de uma pesquisa científica, este último podendo ser tanto uma publicação de um artigo, a apresentação em um evento científico ou a defesa de uma dissertação/tese, mas que, atualmente, pode ser a publicação ou divulgação dos dados coletados/registrados da pesquisa realizada (SAYÃO; SALES, 2016b; TORINO; ROA-MARTINEZ; VIDOTTI, 2020).

Há muito tempo que se reconhece a necessidade de coletar, organizar e acessar os dados ou registros científicos para que pesquisadores desenvolvam, corrijam ou deem continuidade aos seus trabalhos. Um exemplo, lembrado por Sagan (1982), é o caso de Kepler, astrônomo do século XVI, que necessitava de registros astronômicos para corrigir sua teoria do sistema solar (ele acreditava que o sistema solar possuía uma trajetória circular, ao invés de elíptica).

Somente outro astrônomo, o pesquisador dinamarquês Tycho, possuía tais registros, fazendo com que Kepler precisasse viajar em busca desses dados. Sagan (1982, p. 59) afirmou que “[...] Tycho foi o maior gênio observacional da época, e Kepler o maior teórico. Cada um sabia que, sozinho, seria incapaz de atingir uma síntese de um sistema de mundos preciso e coerente, que ambos percebiam iminente”. Essa breve passagem mostra como só o registro detalhado ou uma abordagem teórica de uma pesquisa talvez não seja o suficiente para o processo de pesquisa científica, mas também a importância de se ter acesso aos dados confiáveis e de qualidade.

Com o tempo, e o surgimento de novas necessidades no campo da pesquisa e comunicação científicas, outros recursos e tecnologias foram sendo desenvolvidos e empregados para facilitar as publicações e o acesso aos resultados das pesquisas, como a criação de periódicos científicos ou uso de resumos, o que demandou a participação de diversos profissionais nos ciclos de desenvolvimento e publicação de uma pesquisa.

Burke (2002) apontou a importância do papel dos bibliotecários e arquivistas em uma nova realidade informacional, em razão do grande número de publicações

que surgiam. Nessa perspectiva, para Freire (2006, p. 8), “[...] os profissionais de informação vão estar sempre próximos dos cientistas, seja na organização e na armazenagem das informações produzidas, seja na sua comunicação”. Assim, as atividades de se organizar e armazenar publicações científicas, além do compartilhamento destes estudos, mostram como pesquisas sobre organização e representação da informação e do conhecimento e comunicação científica se complementam.

Paul Otlet, em seu clássico *Tratado de Documentação*, publicado em 1934, já havia refletido e exposto a importância da documentação para estas atividades, especialmente quando afirmou acerca da necessidade de se “[...] analisar, generalizar, ordenar e sintetizar dados adquiridos nas áreas do livro”, além de se “[...] elaborar uma série completa de ‘formas documentárias’ [grifo do autor] para onde possam afluir os dados do pensamento científico ou da prática, desde o simples documento até os complexos das grandes coleções” (OTLET, 2018, p. 12).

Isso faz com que se reflita a respeito do papel dos profissionais de informação no início do século XXI, quando estes participam ativamente das discussões, das atividades e das pesquisas voltadas para o armazenamento, a organização, a representação, a recuperação e o acesso à informação, explorando os ambientes digitais, indo além dos trabalhos tradicionais com documentos bibliográficos, auxiliando na atividade da investigação científica e processos que utilizam os resultados de uma pesquisa, entre outros conjuntos de informação (textos, imagens, planilhas).

Oliveira (2019, p. 104) destaca que a forte presença da modalidade digital na pesquisa científica influenciou na “geração, gestão e disseminação das publicações científicas”, que passaram a envolver grandes volumes de dados, sendo necessário, além de diferentes formas para coletar, gerar e manipular os dados pelo pesquisador, buscar diversas maneiras e instrumentos para o profissional da informação descrevê-los e disponibilizá-los.

Para isso, segundo Hjørland (2013) e Ibekwe-Sanjuán e Bowker (2017), será necessário ir além dos clássicos esquemas de classificação bibliográfica, tesauro e, inclusive, códigos de catalogação. Sendo urgente também uma investigação sobre os recursos digitais que podem auxiliar nessas atividades, como o uso de *Linked Data* (dados conectados), padrões de metadados interoperáveis, entre outros.

Esses recursos podem ser empregados em um Repositório Digital Integrado,

proposto por Ribeiro (2019), um ambiente digital que busca integrar as publicações científicas e os dados de pesquisa, criando um conjunto documental, possibilitando maior acesso e transparência à atividade de comunicação científica.

1.1 PROBLEMATIZAÇÃO

Assim como a Ciência da Informação se institucionalizou e expandiu seu espaço de atuação com o desenvolvimento científico e tecnológico no Pós-Segunda Guerra Mundial, com a expansão do uso de tecnologias digitais no campo científico, na atualidade, os profissionais de informação se depararam com um novo desafio: assegurar o acesso aos dados científicos das pesquisas, incluindo o acesso aos resultados negativos das pesquisas realizadas (SAYÃO; SALES, 2020), como problemas na coleta de dados, resultados inesperados, entre outros.

Se antes o desafio era o acesso a uma determinada publicação, seja por obstáculo geográfico ou por limitações de materiais bibliográficos, como lembrou Cortês (2006), resultando no surgimento dos primeiros periódicos científicos (MEADOW, 1999), atualmente, com o aumento exponencial de publicações, geração de dados e o uso de tecnologias digitais na pesquisa, os esforços são direcionados aos meios de acesso aos dados científicos.

Sayão e Sales (2016b, p. 91) afirmaram que “[...] dados digitais de pesquisa que muito rapidamente deixam de ser meros subprodutos das atividades de pesquisa e se tornam um foco de grande interesse para todo o mundo científico”, impõem a necessidade de uma nova abordagem na organização e representação desses recursos informacionais, além do próprio contexto onde estes foram produzidos.

Os repositórios de dados científicos surgem como uma das principais fontes de acesso aos dados das pesquisas científicas, porém, como afirmou Burke (2002, p. 173), “[...] parece inevitável que nas atividades humanas todas as soluções de um problema mais cedo ou mais tarde acabem gerando outros problemas”. A constatação feita por Burke pode ser confirmada por Sayão e Sales (2016a, p. 68) ao afirmarem que:

[...] o uso e a geração intensiva de dados pelas atividades acadêmicas e de pesquisa criam a necessidade urgente de infraestruturas gerenciais e tecnológicas que tratem de forma dinâmica o ciclo de vida dos dados – do seu planejamento até a seu arquivamento confiável – e insira esses ativos na infraestrutura mundial de informação para a pesquisa.

As infraestruturas gerenciais e tecnológicas auxiliam, também, no processo de representação, organização e comunicação da informação científica, ainda mais na necessidade de se descrever um conjunto documental de uma pesquisa, fazendo o uso de padrões de metadados interoperáveis, além da análise de domínio e terminologias de especialidade.

Porém, como expôs Brooks (2001), no período da popularização da internet, a descrição de informações digitais já enfrentava adversidades em razão do rápido crescimento das informações publicadas e da descentralização desses processos de descrição, representação e arquivamento das informações.

Nesta discussão sobre a descentralização, surgem propostas de repositórios digitais integrados, onde poderá se recuperar os conjuntos de dados depositados e as publicações, permitindo a devida contextualização dos referidos dados. Assim, necessita-se de diretrizes e técnicas capazes de garantir a recuperação e o reuso de dados de pesquisas em um ambiente integrado, chamado de Repositório Digital Integrado (RIBEIRO, 2019).

Este Repositório Digital Integrado poderá auxiliar no desafio de localizar e ter acesso tanto aos dados de pesquisa quanto à contextualização destes dados em suas publicações. Nesse sentido, há a premência em se discutir sobre a organização e representação dos dados de pesquisas em conjunção às publicações científicas, chamadas de conjunto documental (RIBEIRO, 2019).

Portanto, indaga-se a respeito dos instrumentos e recursos utilizados em repositórios digitais integrados indexados nas bases de dados R3data, *Registry of Open Access Repositories* (ROAR) e *Directory of Open Access Repositories* (OPENDOAR). Também se questiona: quais elementos da organização e representação da informação são empregados no conjunto de metadados para descrever o conjunto documental disponível nos repositórios digitais integrados?

1.2 JUSTIFICATIVA

Com o crescimento do uso de repositórios digitais para o armazenamento de publicações e dados de pesquisas, os registros informacionais nesses ambientes digitais necessitam de consistência e clareza, especialmente quando se pensa em interoperabilidade entre registros de diferentes repositórios e a reutilização de dados de pesquisa. Com isso, a importância do uso de padrões de metadados interoperáveis

explicita a necessidade de se pesquisar formas para se aprimorar esses registros.

A relevância econômica se fundamenta na necessidade de se facilitar o acesso aos dados de pesquisa já coletados e registrados, permitindo a reutilização dessa informação, com isso pode-se reduzir o tempo e o custo de pesquisas, uma vez que evitaria o retrabalho na coleta, descrição e organização dos mesmos dados de pesquisa.

A relevância científica deste projeto de pesquisa se justifica na necessidade de se pesquisar o desenvolvimento e a utilização de recursos para a representação, organização e recuperação de dados de pesquisa, assim como a sua integração com o resultado da pesquisa, isto é, a sua publicação e disponibilização. Pois, como afirmaram Pampel e outros (2013, p. 1, tradução nossa),

[...] em relação ao propósito do acesso para uso e reuso dos dados de pesquisa, dados digitais de pesquisa não têm valor sem seus metadados e documentação apropriada descrevendo seu contexto e as ferramentas utilizadas para sua criação, armazenamento, adaptação e análise.

A relevância científica também se justifica acerca da reutilização de dados de pesquisa e a facilidade em seu acesso, além de um viés econômico, que permitem a transparência na produção científica, dado que o conjunto documental da pesquisa estará acessível para todos os cidadãos, e não só para a comunidade científica, incluindo diferentes tipos de profissionais, que poderão checar os dados coletados, divulgá-los, reforçando a creditação no fazer científico e garantindo as boas práticas, inclusive no processo de metodologia da pesquisa.

No âmbito pessoal e profissional, este estudo é a continuação de uma pesquisa realizada durante a graduação na área da Biblioteconomia, pois complementa um tema que uniu o domínio da Organização da Informação ao da Comunicação Científica, por meio da análise da organização da informação científica em periódicos científicos eletrônicos.

Na investigação atual, busca-se analisar as formas de organização e representação do conjunto documental em repositórios digitais integrados, onde tem-se a necessidade de agregar outros recursos de organização e representação desta informação, como o uso de esquemas de metadados específicos para certos domínios, o uso de *Linked Data*, para auxiliar na interoperabilidade entre recursos, além do uso de outros instrumentos para a organização da informação que possam ser aplicados ao repositório digital integrado.

1.3 OBJETIVOS

Nesta subseção, serão apresentados o objetivo geral e os objetivos específicos que conduziram esta pesquisa.

1.3.1 Objetivo Geral

- Analisar o uso de recursos instrumentais para organização e representação do conjunto documental em repositório digital integrado.

1.3.2 Objetivos Específicos

- Identificar as características do repositório digital integrado;
- Verificar as características do conjunto documental;
- Investigar o uso de esquemas de metadados para a organização e representação do conjunto documental.

1.4 SISTEMATIZAÇÃO METODOLÓGICA DA PESQUISA

Quadro 1 – Sistematização metodológica da pesquisa

Problema	Quais elementos da organização e representação da informação são empregados no conjunto de metadados para descrever o conjunto documental disponível nos repositórios digitais integrados?
Objetivo Geral	Analisar os recursos instrumentais para organização e representação do conjunto documental em repositório digital integrado.
Objetivo Específico 1	Identificar as características do repositório digital integrado.
Objetivo Específico 2	Verificar as características do conjunto documental.
Objetivo Específico 3	Investigar o uso de esquemas de metadados para a organização e representação do conjunto documental.

Procedimentos Metodológicos	Pesquisa de natureza descritiva, utilizando-se de levantamento bibliográfico e documental.
------------------------------------	--

Fonte: Elaborado pelo autor.

1.5 ESTRUTURA DA PESQUISA

A pesquisa encontra-se organizada em seções e subseções concernentes à Organização e Representação da Informação, que discute os conceitos pertinentes ao domínio, também apresenta alguns recursos discutidos na literatura da área, como instrumentos para representação da informação em ambiente digital, *Linked Data*, dados abertos, esquemas de metadados, identificadores persistentes e interoperabilidade; à Comunicação Científica, discorrendo sobre a Ciência Aberta, dados de pesquisa e repositórios de dados de pesquisa e integrados; e aos procedimentos metodológicos, descrevendo as cinco etapas da coleta de dados e análise dos resultados. Iniciando-se a discussão, na próxima seção, apresenta-se o tema sobre Organização e Representação da Informação.

2 ORGANIZAÇÃO E REPRESENTAÇÃO DA INFORMAÇÃO

Aborda-se, nesta seção, a fundamentação teórica e a prática que embasou esta pesquisa relacionada à temática sobre a organização e representação do conjunto documental, incluindo a exploração do tema organização da informação nos ambientes digitais, além do uso de recursos para este propósito, como *Linked Data*, padrão de metadados e a importância da interoperabilidade para se realizar a organização e a representação da informação.

2.1 ORGANIZAÇÃO E REPRESENTAÇÃO DA INFORMAÇÃO: CONCEITUAÇÃO

Organizar é uma atividade essencialmente humana. Nesse sentido, Café e Sales (2010, p. 117) lembram que “precisamos organizar para poder compreender o mundo e nos comunicarmos melhor” e Souza (1998) completa ressaltando que organizamos para identificar e tornar algo acessível. Essas afirmações são precisas, especialmente quando se trata de comunicação científica, divulgação e compartilhamento de dados de pesquisa, pois quanto melhor organizado um conjunto de informações científicas, mais fácil para este conjunto ser descoberto por um usuário.

Segundo Café e Sales (2010, p. 118), em um sistema de informação, “[...] a organização da informação é um processo de arranjo de acervos tradicionais ou eletrônicos realizado por meio da descrição física e de conteúdo (assunto) de seus objetos informacionais”. No caso desta pesquisa, trata-se de descrever e organizar as coleções de dados de pesquisa.

Lima (2020, p. 63) também afirma que a organização da informação “compreende processos de representações de objetos informacionais”, cujo objetivo é a recuperação eficaz por parte dos usuários, sendo a representação o produto desses processos, em que a informação é “entendida como um conjunto de elementos descritivos que representam os atributos de um objeto informacional específico.” (BRÄSCHER; CAFÉ, 2008, p. 5).

Trabalhos clássicos, como o de Bräscher e Café (2008, p. 6), diferenciam os conceitos de Organização da Informação e de Organização do Conhecimento, sendo que o primeiro refere-se à organização de objetos informacionais, à descrição física e temática, aos acessos aos registros e conteúdo destes objetos, organizando estes

objetos em coleções, já o último trata da organização de conceitos e “construção de modelos de mundo”.

Portanto, concordando com Lima (2020, p. 63), a

[...] Organização do Conhecimento produz representações sistematizadas de conhecimento, que podem ser utilizadas na Organização da Informação, para padronizar as representações dos conteúdos dos recursos informacionais, visando facilitar a recuperação e o uso deles. Entende-se que a organização do conhecimento é condição necessária para a organização da informação, ou mais especificamente dos recursos informacionais, sob o aspecto dos seus conteúdos.

Monteiro (2008) apresentou uma pesquisa sobre o processo de organização da informação em ambientes digitais, auxiliando como um ponto de partida para descobrir novas técnicas e instrumentos para facilitar o acesso às informações em repositórios digitais. Nessa perspectiva, as reflexões de Santos, Neves e Souza (2019) também contribuíram para se discutir outras perspectivas para a Organização e representação da informação.

Maimone, Silveira e Tálamo (2011) discorreram sobre a importância do uso integrado dos processos de representação temática e descritiva dos recursos informacionais, com foco nos modelos FRBR, FRAD e FRSAD. Cabe ressaltar que a IFLA desenvolveu um modelo conceitual de integração de recursos bibliográficos, culminando no documento sobre o *Library Reference Model* (LRM), sob a coordenação de Riva, Le Bœuf e Žumer (2017).

Em suma, como lembrado por Lourenço, Zafalon e Lopes (2020, p. 88), a prática ou a pesquisa em organização e representação da informação

[...] são as características intelectuais que irão permitir que se identifique os tópicos de interesse contidos em um documento, seu público-alvo específico e seu valor informacional para os usuários de informação; as características do suporte documental dão condições de localizar o documento em que tais características intelectuais estão contidas, de modo que seja possível ter acesso ao conhecimento nele registrado.

Importante destacar o que Lima (2020) expôs em sua reflexão sobre a teoria e técnica na organização e representação do conhecimento e da informação em um ambiente *web*, capaz de explorar as diversas potencialidades dos recursos digitais para a representação e organização da informação em ambiente digital, o uso de dados conectados, capazes de serem interpretados por máquinas, além do refinamento dos conjuntos de metadados por meio da combinação de diferentes esquemas de metadados para domínios específicos.

2.2 RECURSOS PARA REPRESENTAÇÃO DA INFORMAÇÃO EM AMBIENTES DIGITAIS

Lima (2020) apresenta alguns recursos que podem auxiliar na organização e representação do conhecimento e da informação, sendo que o conhecimento pode ser representado por meio de Sistemas de Organização do Conhecimento (SOC), como taxonomias, ontologias, sistemas de classificação e tesouros, e a informação representada por meio de metadados, que podem empregar as regras de catalogação (AACR2 e RDA), padrão Dublin Core (entre outros padrões de metadados) e também modelos conceituais para representação da informação, baseados em uma estrutura em *Resource Description Framework* (RDF).

Cabe lembrar, ainda, as reflexões propostas por Silveira (2017), no que concerne ao uso de instrumentos para a representação descritiva, em que a autora destaca o uso prático desses instrumentos para a construção da representação descritiva, sendo o produto desta atividade a representação propriamente dita, a qual, neste estudo, foi a criação da representação do conjunto documental das publicações científicas.

A representação do conhecimento e da informação pode auxiliar a *Web Semântica*, a partir da estruturação do *Linked Data* (RÍOS-HILARIO; MARTÍN-CAMPO; FERRERAS-FERNÁNDEZ, 2012; LIMA, 2020), realizando a conexão entre recursos e melhorando a interoperabilidade entre estes, a partir do uso de padrões de metadados. Também pode-se utilizar a prática de dados abertos, isto é, de livre acesso, compartilhamento e manipulação em conjunção ao uso de Dados Conectados (*Linked Data*) (ISOTANI; BITTENCOURT, 2015).

Assim, serão discutidos alguns recursos e instrumentos acima citados que já são utilizados em repositórios digitais ou debatidos em trabalhos publicados e que poderão ser empregados para a organização e representação do conjunto documental.

2.2.1 Dados abertos conectados

Além da necessidade de se lidar com grandes quantidades de dados, um dos desafios do profissional da informação é trabalhar com as diversas conexões existentes entre os dados, inclusive a semântica (ISOTANI; BITTENCOURT, 2015). Portanto, para uma melhor eficiência, necessita-se da utilização não só de dados

abertos de qualidade, mas que estes estejam também conectados.

Em uma tradução livre, a definição de aberto significa ser livremente acessado, usado, modificado e compartilhado para qualquer propósito (OPEN DEFINITION, 2021), devendo-se, no máximo, se preocupar com requisitos que preservem a proveniência e manutenção das características para serem abertos. Cabe um destaque para os requisitos que garantam a proveniência, característica fundamental para o desenvolvimento de um conjunto documental que foi explorado ao longo deste trabalho.

Tem-se, portanto, a identificação de algumas características que podem ser também exploradas nos dados abertos, como a sua acessibilidade (também deve estar sempre disponível), permissão para seu reuso e compartilhamento, além de permanecerem sempre abertos, isto é, independentemente de quem use.

Os dados conectados possuem uma característica que foi definida (porém sem a possibilidade de ser implementada devido às restrições tecnológicas à época) no próprio advento da internet, ou seja, a possibilidade de tanto os humanos quanto as máquinas poderem interpretar as relações entre os dados.

Os dados conectados utilizam uma estrutura baseada em um padrão formado por *Uniform Resource Identifier* (URI), que pode ser um endereço na *web* ou um outro identificador único, e *Hypertext Transfer Protocol* (HTTP) (igual à *web* de documentos), além do emprego de RDF como modelo de representação da informação e uma linguagem de consulta, chamada de linguagem SPARQL. O RDF é uma linguagem de representação de informações na *web*, o que permite a descrição formal de recursos e que esta seja acessível e interpretada também por máquinas. (ISOTANI; BITTENCOURT, 2015).

O modelo RDF foi desenvolvido pelo W3C, sendo constituído por três objetos: recursos, propriedades e declarações. De acordo com Santarém Segundo (2010), os recursos correspondem a variadas informações que podem ser encontradas em uma página na *web*, no conteúdo de um livro ou CD, em uma pessoa, em um lugar ou em qualquer documento disponível em um repositório ou biblioteca digital, sendo que estas informações podem ser identificadas por uma URI.

As propriedades representam as características do recurso por meio de atributos que distinguem um recurso de outro, podendo também descrever um relacionamento entre atributos. Enquanto as declarações são as informações completas sobre o recurso, com suas propriedades e valores para os campos da

propriedade.

O RDF pode ser utilizado em conjunto com uma outra estrutura, chamada de RDF-Schema, “[...] que dá à linguagem RDF o poder de construção de estruturas como hierarquias, propriedades e subpropriedades, entre outros.” (SANTARÉM SEGUNDO, 2010, p. 118). Assim, como será apresentado na subseção de esquemas de metadados, pode-se aproveitar das funcionalidades dos esquemas de metadados, princípios e práticas para construir esquemas de descrição mais apropriados para determinados domínios.

Porém, os dados conectados não são somente sobre o uso de recursos digitais, mas também se relacionam ao emprego de boas práticas que permitem a publicação e a conexão de dados em um ambiente digital que possibilite que tanto os usuários quanto as máquinas possam interpretar essas conexões. Embora os dados conectados não precisem estar exatamente abertos, para as boas práticas na Ciência Aberta, além de fomentar e facilitar as pesquisas científicas, recomenda-se que dados de pesquisa e as publicações estejam em plataformas abertas, com livre acesso e fácil recuperação.

Assim, seguem as recomendações para o Sistema de 5 Estrelas (ISOTANI; BITTENCOURT, 2015) para os dados abertos: a primeira recomendação, e primeira estrela, refere-se ao uso de formatos com licença aberta, como um arquivo PDF ou JPEG; a segunda recomendação, e segunda estrela, refere-se aos dados publicados de forma estruturada e legível por computador, como em um formato XLS ou XML; a terceira recomendação, e terceira estrela, refere-se à publicação de forma estruturada, porém não em um formato proprietário, como em CSV; a quarta recomendação, e quarta estrela, refere-se à publicação de dados seguindo regras da W3C no formato RDF; a quinta, e última estrela, refere-se ao uso de dados conectados com outros dados, utilizando identificadores persistentes, que permitem a contextualização da publicação.

Em suma, o *Linked Data* é a proposta de uma ligação entre recursos informacionais por meio da utilização de *links* semânticos, sendo esta ligação por meio de uma URI e RDF em uma página da *web*. (SIMIONATO, 2017).

Pode-se, então, começar a depreender que, em um repositório digital integrado, o uso de recursos e práticas que facilitem a conexão entre dados de pesquisa e publicações científicas serão de essencial importância, pois, conforme afirmaram Isotani e Bittencourt (2015, online), “a *Web* de Dados cria inúmeras

oportunidades para a integração semântica dos próprios dados, motivando o desenvolvimento de novos tipos de aplicações e ferramentas, como navegadores e motores de busca”.

2.2.2 Os esquemas de metadados

O uso de um padrão de metadados, ou mesmo mais de um, como bem lembrado por Ríos-Hilario, Martín-Campo e Ferreras-Fernández (2012), tem extrema importância na atividade de organização e representação da informação, assim como na interoperabilidade entre recursos e sistemas. O metadado não é um recurso novo para o profissional da informação, pois, como destacou Caplan (2003), há séculos se utiliza esse recurso em ambiente físico ou analógico, como na catalogação, por exemplo, para se criar padrões de informação em um catálogo.

Lima, Santos e Santarém Segundo (2016, p. 52) lembram que os metadados são cruciais para a compreensão do “[...] recurso armazenado, pois descrevem informações semânticas e sintáticas sobre o recurso”. Dessa forma, os padrões de metadados podem ser divididos de acordo com suas funções, tais como: administrativos, descritivos, estruturais, conservação e técnicos e de uso. (GRACIO, 2002; CATARINO, 2009; ALVES; SANTOS, 2013; ANDRADE, 2018).

Quadro 2 – Funções e descrição das funções dos metadados

FUNÇÕES	DESCRIÇÃO DAS FUNÇÕES
Administrativo	Gestão e administração dos recursos de informação
Descritivo	Descrição, identificação e representação dos recursos de informação
Estrutural	Organização dos recursos de informação
Preservação	Conservação e preservação dos recursos de informação
Técnicos	Funcionamento e comportamento do sistema e metadados
Uso	Nível e tipo de uso dos recursos de informação

Fonte: Adaptado de Catarino (2009), Alves e Santos (2013)

As funções descritas acima não são utilizadas separadamente, sendo importantes ao se analisar o conjunto documental. Os metadados também podem ser utilizados em complementariedade com a documentação em um ambiente digital. Conforme afirmam Crosas e outros (2015, p. 265),

[...] os metadados e a documentação adicional facilitam a descoberta de dados através de ferramentas de busca e informam outros pesquisadores quanto ao formato, variáveis, fonte, metodologia e a análise aplicada aos dados.

Ampliam-se, assim, as funcionalidades para se realizar a recuperação da informação por meio da integração de metadados, inclusive em relação à proveniência do recurso, que na sua ausência pode limitar os resultados da análise do referido recurso. Borgman (2015, p. 76) discorre sobre essa questão ao ressaltar que “[...] quanto mais distante o pesquisador está da fonte original dos dados, maior será a necessidade de informações sobre proveniência para a reutilização dos dados”, caso contrário, o pesquisador deverá confiar na autenticidade e integralidade das informações disponíveis.

Alves e Santos (2013) também destacam que, além das funções dos metadados, tem-se os princípios e práticas para tornarem os metadados mais efetivos. Os princípios de modularidade, extensibilidade, refinamento e multilinguismo são assim definidos:

- A **modularidade** refere-se à capacidade de construir metadados em blocos, camadas ou módulos, permitindo a complementariedade entre diferentes esquemas de metadados, auxiliando inclusive na interoperabilidade entre diferentes conjuntos, podendo, ainda, criar um diferente esquema de metadados por meio da conjunção de diferentes campos.
- A **extensibilidade** indica a capacidade de incluir novos metadados a um determinado conjunto já existente.
- O **refinamento** refere-se à capacidade de atingir uma determinada especificação no conjunto de metadados de acordo com a necessidade para se descrever um objeto ou necessidade de uma determinada comunidade de usuários.
- O **multilinguismo** refere-se a uma necessidade de comunicação entre diferentes grupos culturais ou linguísticos, refletindo a múltipla existência de diferentes formas de se representar um recurso informacional.

Nessa perspectiva, Alves e Santos (2013) apresentam, ainda, as práticas, referentes às regras, restrições e questões de infraestrutura, as quais possuem relação com questões de criação e gerenciamento. Estas práticas são: perfil de aplicação, sintaxe e semântica, associações entre modelos, identificação e

designação de elementos de metadados, registros de metadados, integralidade na descrição, elementos obrigatórios e opcionais, metadados subjetivos e objetivos, geração automática de metadados. São descritas da seguinte maneira:

- O **perfil de aplicação** permite a união dos princípios de modularidade e extensibilidade, combinando diferentes esquemas de metadados para adaptar o conjunto de metadados a uma determinada exigência. O objetivo principal dessa prática é a interoperabilidade semântica.
- A **sintaxe e semântica** referem-se, respectivamente, à forma como os campos dos esquemas de metadados são expressos e ao significado que cada campo possui, facilitando a interoperabilidade entre diferentes domínios ou repositórios.
- As **associações entre modelos** refletem o relacionamento entre diferentes registros de metadados em um mesmo recurso informacional.
- A **identificação e designação de elementos de metadados** representa a forma como cada elemento (ou campo) será expresso em um determinado conjunto de metadados.
- Os **registros de metadados** representam uma coleção de metadados e as definições de seus elementos e valores possíveis. Cria-se um registro de “[...] conjuntos de elementos de metadados, esquemas, perfis de aplicação, esquemas de codificação[...]” (ALVES; SANTOS, 2013, p. 60-61), entre outros.
- A **integralidade na descrição** significa a quantidade de informações que estarão presentes em um conjunto de metadados, os níveis de granularidade. Representa maior ou menor especificidade na descrição de um recurso, suas relações com outros recursos. Depende de diretrizes e políticas estabelecidas para a descrição do recurso.
- Os **elementos obrigatórios e opcionais** representam quais elementos são essenciais para descrição de um recurso e quais são livres para serem utilizados. Depende do tipo de recurso, do domínio e da comunidade que irá utilizar. Também há a relação com políticas do domínio, para existir uma padronização, visando a interoperabilidade.
- Os **metadados objetivos e subjetivos** são aqueles, respectivamente, que possuem características de fato, como título ou autoria, e aqueles

que são atribuídos, como resumo e palavras-chave/descriptores. Nesse caso, é necessário explicitar o contexto do recurso, permitindo também a interpretação por máquinas/aplicativos.

- A **geração automática de metadados** permite uma pré-configuração para a extração de informações específicas do recurso, mas também pode ser combinado com a geração manual de metadados, auxiliando na redução dos custos e tempo de trabalho.

As autoras Alves e Santos (2013) apontam que podem existir outros princípios e práticas, a depender das necessidades do domínio, do tipo de recurso arquivado e dos usuários do sistema de informação, porém os elementos acima citados são alguns dos mais gerais apresentados, auxiliando na identificação de recursos a ser empregados para a descrição dos metadados para o conjunto documental.

Cabe um destaque aos metadados que auxiliam no controle de vocabulário, especialmente aqueles que permitem o relacionamento entre termos em uma lista estruturada (FUJITA; TOLARE, 2019), utilizando-se dos recursos para dados abertos e conectados (RÍOS HILARIO; GIL URDICIAIN, 2015). Como exemplos, tem-se o uso do SKOS, apresentado por Sousa e Ramalho (2020), e do *Metadata Authority Description Schema* (MADS), para implementação da *Library of Congress Subject Headings* (LCSH) (RÍOS HILARIO; GIL URDICIAIN, 2015), utilizando-se da estrutura de RDF, possibilitando a vinculação de um conceito a um vocabulário controlado eletrônico, também construído em RDF, por meio de uma URI.

Atualmente, existe uma grande variedade de esquemas (ou padrões) de metadados, podendo ser para propósitos mais gerais ou para domínios específicos. O padrão Dublin Core é um dos mais utilizados para os propósitos gerais, por sua flexibilidade e simplicidade. Para domínio específico, pode-se ter como exemplo:

- **Metadados para dados de pesquisa:** *Data Catalog Vocabulary* (DCTA), recomendado pelo W3C, que apresenta um vocabulário (controlado) em RDF, o que facilita a interoperabilidade entre dados catalogados no ambiente *web*. O *Data Documentation Initiative* (DDI) é um padrão internacional que descreve os dados produzidos por questionários e outros métodos de observação nas áreas das Ciências Sociais, comportamentais, Economia e Ciências da Saúde.
- **Metadados para publicação e comunicação de imprensa:** ONIX, um padrão internacional para representação e comunicação de livros em

formato eletrônico. EPUB, padrão internacional para intercâmbio de publicações e documentos digitais, além de definir uma forma de representar, empacotar e codificar conteúdos *web* aprimorados (*enhanced*) em diversos formatos, como HTML5, CSS, SVG, imagens, entre outros.

Além desses, pode-se apresentar o esquema de metadados com propriedades específicas para a identificação consistente de um recurso para a sua citação, o *DataCite*. O grupo de trabalho do *DataCite* propõe estreitar laços com outras comunidades de pesquisadores, como os responsáveis pelo *Open Researcher and Contributor* (ORCID), *Digital Object Identifier* (DOI) e do próprio *Dublin Core*. (DATA CITE..., 201-).

Segundo Príncipe e Schirrwagen (2015), a *Open Access Infrastructure for Research in Europe* (OpenAIRE) recomenda em suas diretrizes a utilização dos esquemas *Dublin Core*, para repositórios institucionais (bibliográficos), e o *DataCite*, para os repositórios de dados. Podendo explorar o princípio da modularidade para se combinar os esquemas *Dublin Core* e *DataCite* em repositórios digitais integrados.

2.2.3 Identificadores persistentes de autoria

A identificação de autoria é, também, um importante recurso na descrição dos recursos informacionais, existindo ainda a possibilidade do uso de dados conectados para sua identificação única e persistente. Serra e Andrade (2018) apresentaram algumas possibilidades, tanto em nível internacional, como o ORCID, o *ResearcherID* e o *Virtual International Authority File* (VIAF), quanto em nível nacional, como a identificação via plataforma Lattes.

Segundo estas autoras, o Lattes¹, o *ResearcherID*² e o ORCID³ têm como características a descrição realizada pelos próprios autores, podendo gerar ambiguidades e falta de padronização, variações nas grafias dos nomes e complementações sobre a afiliação, áreas de interesse, entre outros. Enquanto o VIAF⁴ é produzido por um serviço de informação, a *Online Computer Library Center* (OCLC).

¹ Plataforma Lattes: <https://lattes.cnpq.br/>

² ResearchID: <https://www.researcherid.com/>

³ ORCID: <https://orcid.org/>

⁴ VIAF: <https://viaf.org/>

O Lattes é mantido pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), reunindo em um único lugar informações sobre o currículo dos pesquisadores, os grupos de pesquisa e as instituições das quais fazem parte. As informações no Lattes são integradas a outras bases de dados, como a SciELO, LILACS, Scopus e ORCID (SERRA; ANDRADE, 2018). Os dados são, além de conectados, abertos, permitindo a utilização da ferramenta *Lattes Extrator*.

O *ResearcherID* é uma plataforma onde o pesquisador pode criar um perfil de forma gratuita, gerando um identificador único de 11 a 13 caracteres, sendo integrado à plataforma da *Web of Science* e, também, ao ORCID. Esse identificador permite que o pesquisador registre suas produções, diferencie-se de pesquisadores com nomes parecidos e que se relacione com potenciais colaboradores acadêmicos.

O ORCID também é uma plataforma digital que gera um identificador único com 16 caracteres em formato HTTP, isto é, um *Uniform Resource Locator* (URL). O perfil é de criação gratuita, porém pode haver cobrança para perfis institucionais com vínculos a repositórios, revistas ou outras plataformas (SERRA; ANDRADE, 2018). O perfil permite a inclusão de informações sobre dados biográficos, produções, afiliação, financiamento, entre outros.

O VIAF é uma iniciativa desenvolvida e mantida pela OCLC, sendo que os registros são criados e catalogados pelas Bibliotecas Nacionais e outras agências catalogadoras. Segundo Salgado (2005), o VIAF tem como primeiro objetivo a criação de um *link* (dado conectado) entre os catálogos de autoridade das Bibliotecas Nacionais. Portanto, o VIAF é uma iniciativa que não visa somente a identificação no meio acadêmico, mas um identificador único com os nomes preferidos e suas variações em uma escala geral e global de produção.

Assim, os recursos para identificação de autores via identificadores únicos é mais um elemento para a interoperabilidade entre as produções científicas, sendo este o assunto a ser discutido na próxima subseção.

2.2.4 Interoperabilidade

A interoperabilidade é a capacidade de diferentes sistemas de informação comunicarem-se entre si, permitindo a troca de serviços e dados com perda mínima de conteúdo. Para isso, necessita-se de um entendimento em comum entre os provedores e as requisições solicitadas.

O uso de padrões de metadados facilita a interoperabilidade, que possui uma variedade de definições, aspectos e níveis, segundo o levantamento na literatura realizado por Zeng (2019) e Cunha (2021). No levantamento de Cunha (2021), a interoperabilidade tem aspectos que podem ser identificados como: técnico, estrutural, de conteúdo, semântico, organizacional, internacional, sintático, baseado em serviços, político/humana, baseado no usuário, legal e intercomunitário.

A autora supracitada destaca seis aspectos da interoperabilidade, apresentados no quadro a seguir.

Quadro 3 – Aspectos da interoperabilidade

ASPECTOS	DESCRIÇÃO DOS ASPECTOS
Técnico	Relacionado ao uso de padrões e protocolos para assegurar a interoperabilidade
Semântico	Relacionado ao significado das informações e uso de ferramentas para representação da informação
Político/humano	Relacionado à maior acessibilidade dos recursos de informação
Intercomunitário	Relacionado à informação interdisciplinar e de diferentes áreas e sistemas
Legal	Relacionado aos requisitos legais de acesso à informação
Internacional	Relacionado às diferenças de idiomas, comunicação, culturais, entre outros

Fonte: Adaptado de Cunha (2021)

Cunha (2021) apresenta também os níveis de interoperabilidade, sendo:

- **Federação:** o nível mais alto de interoperabilidade, com a utilização dos mesmos protocolos e padrões para maior integração entre os entes federados.
- **Harvesting:** consiste na coleta dos metadados dos provedores (serviço ou dados), permitindo a utilização de padrões de metadados diferentes entre os sistemas, desde que compreensível por um protocolo em comum. O protocolo mais utilizado é o OAI-PMH.
- **Agregação:** na ausência de padrões e protocolos em comum, utilizam-

se níveis básicos de comunicação, possibilitando o uso de metabuscadores.

Concluindo esta seção, percebe-se que as funcionalidades do *Linked Data* (dados conectados), os aspectos dos padrões de metadados (em especial a modularidade e extensibilidade entre esquemas de metadados), os identificadores persistentes para autoria e interoperabilidade auxiliarão em uma melhor comunicação científica, facilitando as diversas práticas da Ciência Aberta, como será visto na seção a seguir.

3 COMUNICAÇÃO CIENTÍFICA: DOS PERIÓDICOS CIENTÍFICOS AOS DADOS DE PESQUISA

Nesta seção, será discutida a temática da Comunicação Científica, incluindo o movimento de Ciência Aberta e os conceitos de dados de pesquisa, de repositório de dados de pesquisa, de repositórios digitais integrados e do conjunto documental de uma pesquisa.

3.1 COMUNICAÇÃO CIENTÍFICA: ORIGENS E DISCUSSÕES

A comunicação científica, conforme afirma Meadows (1999), não pode ser identificada acerca de seu surgimento, mas, para o autor, a Grécia antiga causou um impacto significativo a esse processo. A atividade e os resultados de uma pesquisa poderiam ser comunicados tanto por via escrita quanto por via oral, e os gregos antigos utilizavam ambas. Pela via oral, havia as discussões na Academia ou os simpósios, enquanto registrava-se em manuscritos, por exemplo, os ensinamentos de Aristóteles.

Garvey (1979) define a comunicação científica como uma atividade de troca de informações que ocorre, geralmente, entre cientistas ativamente envolvidos em uma pesquisa, incluindo uma discussão informal entre dois pesquisadores até a formalização dos resultados ou o andamento de uma pesquisa em uma publicação periódica, uma revisão ou um livro. A formalização da comunicação científica significou uma importante realização para a atividade de pesquisa.

Uma publicação informal tem um caráter mais efêmero e restrito a um grupo, enquanto uma publicação formal, em um artigo ou livro, permite a disponibilidade para um maior grupo de pessoas por mais tempo (MEADOWS, 1999), lembrando que a comunicação informal e a formal se complementam.

Enquanto os periódicos científicos contribuíram para a formalização da comunicação científica ao longo do tempo, no século XX, como destacou Muller (2006), o evento mais importante para a comunicação científica foi o surgimento do Movimento de Acesso Livre.

O desenvolvimento tecnológico na área da informática possibilitou o surgimento dos periódicos científicos eletrônicos, causando mudanças no fluxo informacional e resultando “[...] na descentralização do controle e na alteração da

cadeia de produção, disseminação e utilização do conhecimento científico registrado, antes exclusivo dos editores comerciais de revistas científicas.” (WEITZEL, 2005, p. 162).

Com isso, dá-se início ao Movimento de Acesso Aberto, ou *Open Archives Initiative* (OAI), inaugurando um novo modelo de comunicação científica, que visava refletir acerca dos principais fundamentos da ciência e da disseminação da informação científica. Weitzel (2005) indica alguns dos principais elementos do OAI, como: autoarquivamento pelos autores; metadados padronizados para descrição; acesso livre à produção científica e o protocolo OAI.

Ainda assim, como afirmou Oliveira (2019), na contemporaneidade, a comunicação científica era realizada baseada na publicação impressa ou na divulgação dos resultados finais de uma pesquisa, ocorrendo a mudança com o Movimento de Acesso Aberto, que auxiliou no desenvolvimento de uma nova forma do fazer científico e sua comunicação, culminando no surgimento do movimento para a Ciência Aberta.

3.2 CIÊNCIA ABERTA

A Ciência Aberta desponta como uma nova prática institucionalizada de se realizar a pesquisa científica e a sua divulgação, calcada na colaboração entre pesquisadores e na modalidade eletrônica e digital (OLIVEIRA, 2019). O Glossário da Ciência Aberta (FUNDAÇÃO OSWALDO CRUZ, [20--]) a define como um

[...] conceito amplo que abarca diversas práticas do fazer científico, desde o acesso aberto à literatura científica e aos dados abertos de pesquisa, [...] *software* e *hardware* abertos, entre outras. Faz referência a um modelo metodológico alinhado à filosofia da cultura digital, da colaboração e do compartilhamento das práticas e dos conteúdos científicos.

Por mais que este conceito seja apresentado de forma ampla, percebe-se uma noção que fundamenta essa atividade: a atividade colaborativa. Além da transparência na apresentação dos dados, acesso à literatura científica e o uso de *software* e *hardware* livres ou de código aberto. Esta ideia está presente nas afirmações de Kon (2013), Sayão e Sales (2016b) e Silva e Silveira (2019).

Silva e Silveira (2019, p. 2) lembram que “a Ciência Aberta é um movimento que incentiva a transparência da pesquisa científica desde a concepção da investigação até o uso de softwares abertos”, afinal, a transparência da pesquisa

científica ressalta as boas práticas e confiabilidade nos resultados alcançados pelo pesquisador.

Sayão e Sales (2016b) apontam como a transição na comunicação científica vai em direção à disponibilização dos dados de pesquisa, principalmente com o aprimoramento de ferramentas para o uso em ambiente digital, especialmente no que concerne à gestão de dados da pesquisa.

Inclusive, Kon (2013, p. 28) enfatiza como a ideia de Ciência Aberta extrapola o Movimento de Acesso Aberto, que preconizou o acesso livre às publicações, pois “[...] Ciência Aberta não se limita à forma das publicações abertas”, sendo necessário acesso aos “dados e ferramentas utilizados nos experimentos, simulações e análises”. Necessita-se, portanto, de dados abertos e códigos abertos (*software* livre).

Assim, ao abordar o tema Ciência Aberta, foi dada especial ênfase à discussão sobre dados de pesquisa (ou científicos), repositórios de dados de pesquisa e repositórios digitais integrados.

3.3 DADOS DE PESQUISA

O conceito de dados, juntamente com os de informação e conhecimento, é um dos mais discutidos no campo da Ciência da Informação. Dessa maneira, pode ser considerado, em uma definição clássica de Miranda (1999, p. 286), um “[...] conjunto de registros qualitativos ou quantitativos conhecido, que organizado, agrupado, categorizado e padronizado adequadamente transforma-se em informação”. Assim sendo, os dados necessitam ser descritos e representados.

Santos e Sant’anna (2019) ressaltam que os dados têm um esquema em formato tríade (entidade <e>, atributo <a> e valor <v>), sendo que a estrutura do dado é estabelecida pela identificação dos elementos <e> e <a>, entidade e atributo, respectivamente. Os referidos autores ainda descrevem os dados como estruturados e não-estruturados. Nesse sentido, dados estruturados possuem uma “estrutura interna explícita e seu conteúdo <v> interno é evidenciado por meio de uma semântica que permite que todo o seu conteúdo seja interpretado de forma autônoma por máquina.” (SANTOS; SANT’ANNA, 2019, p. 58). Por outro lado, os “[...] dados não estruturados [...] exigem interpretador externo para identificar sua estrutura e semântica e interpretar seu conteúdo”. (SANTOS; SANT’ANNA, 2019, p. 58)

Nessa perspectiva, Sayão e Sales (2016b) indicam que mesmo o termo

“dados de pesquisa” pode variar conforme o domínio do conhecimento, inclusive pode ser dado de um resultado negativo de uma pesquisa (SAYÃO; SALES, 2020), os resultados que não confirmavam uma hipótese ou dados inesperados, além de possíveis problemas no desenvolvimento de uma pesquisa.

Portanto, para se trabalhar com os dados de pesquisa, necessita-se desenvolver um ambiente digital que facilite o armazenamento, a organização e a recuperação desses dados. Dois ambientes podem ser explorados, um repositório de dados de pesquisa e um repositório digital integrado.

3.4 REPOSITÓRIO DE DADOS DE PESQUISA

Um repositório de dados de pesquisa tem como objetivo garantir o acesso contínuo e livre aos dados produzidos por uma pesquisa científica. Como já explicitado anteriormente, necessita de uma infraestrutura gerencial e tecnológica capaz de garantir o ciclo de vida dos dados de pesquisa (SAYÃO; SALES, 2016). Isso requer uma abordagem interdisciplinar, pois cada etapa necessita de conhecimentos específicos.

Sales (2014) e Sayão e Sales (2016) apresentam alguns benefícios e funções que um repositório de dados de pesquisa pode oferecer: visibilidade e compartilhamento dos dados, crédito ao autor e preservação digital, memória científica e transparência, segurança dos dados e disponibilidade, curadoria digital, reuso dos dados, rede de repositório (interoperabilidade), indicação de produtividade, qualidade e outros serviços inovadores.

O ambiente digital de um repositório de dados de pesquisa deve ser dinâmico e flexível, pois lida com uma variedade grande de tipos de dados, podendo estes serem textuais (pdf, docx, txt), estatísticos (csv, xls), imagéticos (jpeg, png), em forma de código de programa. Portanto, cada tipo de arquivo possui seu correspondente em formato aberto ou formato proprietário e isso influenciará em seu uso por parte do usuário ou pesquisador.

Nesse ambiente digital, o profissional da informação deverá superar os esquemas clássicos de organização da informação e do conhecimento para poder oferecer o conteúdo tanto para os pesquisadores quanto para outros usuários que deste necessitarem.

Pampel e outros (2013) e Sayão e Sales (2016) elencam alguns tipos de

repositórios de dados pesquisa, como os repositórios institucionais, administrados por uma instituição acadêmica ou de pesquisa; repositórios multidisciplinares, que abarcam coleções de dados de uma diversidade de domínios do conhecimento, mas não são geridos por instituições; repositórios de dados orientados para projetos, onde os dados estão organizados por projetos de pesquisa ou para solução de problemas específicos.

Assim, Aventurier e Alencar (2016) afirmam que o dado de pesquisa serve, inicialmente, para validar os resultados de uma pesquisa publicada, devendo ser depositado concomitantemente com a publicação científica, por exigência dos editores científicos. Desse modo, ao enviar uma publicação, o pesquisador deve disponibilizar os dados e metadados para os avaliadores.

Este procedimento de unir dados de pesquisa e publicações surge como uma opção para outro tipo de repositório de dados, um que seja integrado em todo o ciclo de pesquisa.

3.5 REPOSITÓRIO DIGITAL INTEGRADO

Uma questão, levantada por Ribeiro (2019), se refere ao início do desenvolvimento de repositórios de dados de pesquisa no Brasil, argumentando que, devido a certo atraso na implantação desses repositórios, as instituições poderiam aproveitar a já exitosa experiência com os repositórios institucionais digitais, pois os repositórios digitais permitem a publicação de diferentes fontes e a integração de metadados por mecanismos de buscas, conforme destacado por Andrade (2018).

Além das funcionalidades apontadas acima, tem-se a infraestrutura já instalada, a alocação de recursos humanos qualificados, as técnicas e recursos para organização e representação da informação. Há, também, um dos principais benefícios para o uso de repositórios digitais integrados que é a facilidade de acesso ao conjunto documental, isto é, a publicação (artigo em periódico, evento, dissertação e tese) e seus dados de pesquisa depositados, aproveitando-se a infraestrutura gerencial e tecnológica existente.

Corroborando as afirmações acima, em um relatório apresentado por Ivanovic e outros (2019), o *software* livre DSpace é utilizado em 42% das instituições europeias pesquisadas, tendo como outros *softwares* o Eprint, Dataverse, Figshare, Pure Plattaform, Diva consortium, Samvera, Invenio e Digitool.

Os profissionais de informação já trabalham, em ambientes digitais, com fluxos informacionais e gestão da informação, porém, cabe identificar e analisar outras formas de organização e representação da informação, especialmente a científica, neste ambiente digital flexível e dinâmico. Uma forma seria integrar os dados de pesquisa e as publicações resultantes das pesquisas, que será discutida como conjunto documental.

3.6 CONJUNTO DOCUMENTAL

A ideia de unir a publicação científica (artigo ou livro) e os dados de pesquisa não é recente, tendo sido discutida no contexto da publicação ampliada (SALES; SOUZA; SAYÃO, 2014; RODRIGUES; SANT'ANA, 2016; ARRAIZA; LIBERATORE; VIDOTTI, 2018).

Sales, Souza e Sayão (2014, p. 3474) evidenciaram a necessidade de se apresentar novas formas de publicações científicas decorrentes de um novo fazer na ciência, que seria decorrência da geração, compartilhamento e uso de dados digitais de pesquisa por meio de recursos digitais, usando-se o conceito de publicação ampliada, que seria “[...] a ligação entre publicações digitais convencionais artigos de periódicos, teses e dissertações, por exemplo - e os dados de pesquisa que dão sustentação aos argumentos do texto”, fazendo-se uso de recursos digitais para a descrição e organização desse conjunto documental.

Nesse sentido, Ribeiro e Sant'ana (2016) tratam a publicação ampliada como a composição de objetos digitais, no próprio repositório ou em um ambiente externo, tais como os documentos, os dados gerados por uma pesquisa, os metadados e os elementos para sua descrição. Este conjunto de objetos digitais podem ser relacionados entre si, por meio de enlace (*links*) criando uma estrutura que lhe permite uma identificação única, com características próprias e que se possa navegar entre os objetos.

Os pesquisadores Arraiza, Liberatore e Vidotti (2018) afirmam que o tratamento para uma publicação ampliada se divide em três abordagens: modular, classes e instâncias e semântica. A publicação modular é um conjunto de informações suficientes para transmitir um significado, podendo ser elementares ou complexos, sendo este último dividido em dois tipos: compostos (agregando módulos independentes) ou aglomerados (centrado em um conceito que vai se expandindo nos

elementos restantes). Os autores destacam que cada módulo deve possuir seus próprios metadados.

A segunda abordagem - classes e instâncias - enfatiza a unidade de comunicação científica, segundo Arraiza, Liberatore e Vidotti (2018), não discriminando os formatos, podendo, assim, reunir as partes que não estejam juntas a partir de identificadores persistentes. Já a terceira abordagem - semântica – utiliza-se de um modelo semântico, composto por duas fases: uma baseada em ontologia e a outra na interface do sistema de submissão, para adicionar as principais afirmações do texto junto com os metadados bibliográficos.

Sobre a relação entre os dados de pesquisa e a publicação científica, Callaghan (2019) destaca a função de informação de suporte que um artigo fornece aos dados de pesquisa, afinal, enquanto este apenas descreve o seu conteúdo, o artigo científico apresenta as características da coleta, discute mais minuciosamente o significado dos dados coletados, entre outras questões acerca da pesquisa realizada.

A partir do entendimento do que seja o conjunto documental, necessita-se identificar alguns recursos que podem auxiliar na organização e na representação deste elemento. Com base nas funções da catalogação apresentadas por Sayão e Sales (2016b), temos a atribuição de metadados (para descrever o conteúdo do conjunto documental, técnicos e administrativos, além dos metadados de preservação, visando garantir a integridade, autenticidade e proveniência), o emprego de um SOC (ontologia, taxonomia ou tesouro), o uso de identificadores persistentes (DOI, *handles*, URI, ORCID), documentação (informações sobre a pesquisa, projeto, metodologia, entre outros) e o uso de dados vinculados ou conectados (*Linked Data*).

Ribeiro (2019) afirma que os princípios FAIR, que norteiam a disponibilidade de dados, podem ser adaptados para o conjunto documental nos repositórios digitais integrados, que seriam disponibilizados seguindo os princípios os quais garantiriam a encontrabilidade, acessibilidade, interoperabilidade e reuso do conjunto documental.

Portanto, a partir destas funções da catalogação acima e dos princípios FAIR adaptados para o conjunto documental por Ribeiro (2019), pode-se começar a delinear a estrutura deste conjunto, adaptado dos princípios FAIR, sendo explicitado da seguinte maneira:

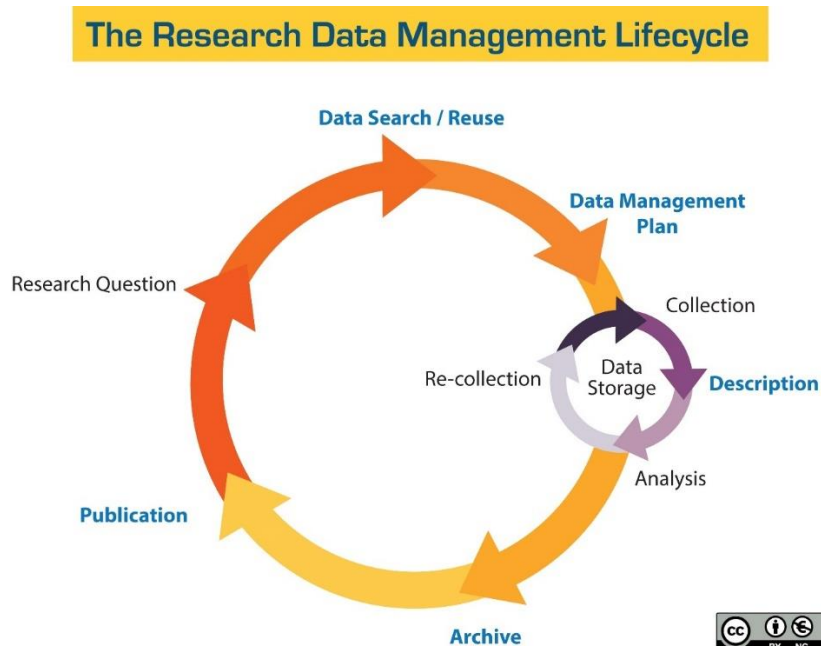
- **Encontrabilidade:** os metadados, os dados e os documentos devem possuir identificadores globais, persistentes e identificáveis; os

metadados devem ser enriquecidos; metadados e documentos devem identificar claramente os dados que descrevem; e todos esses elementos devem estar indexados em uma base que permita a busca.

- **Acessibilidade:** os metadados, os dados e os documentos devem possuir identificadores que permitam a sua recuperação por meio de protocolos de comunicação; o protocolo deve ser aberto, gratuito e universalmente implementável; o protocolo deve permitir a autenticação; e os metadados devem ser acessíveis a qualquer momento.
- **Interoperabilidade:** os metadados, os dados e os documentos devem ser representados por uma linguagem formal, baseados em modelos conceituais; devem utilizar vocabulários compatíveis com os princípios FAIR; devem possuir referências qualificadas.
- **Reuso:** os metadados, os dados e os documentos devem ser descritos com pluralidade de atributos precisos e relevantes; licenças de uso devem estar disponíveis; associados à sua proveniência; e em consonância com os padrões do domínio.

Esses princípios possuem relação direta com o uso de esquemas de metadados apropriados para a tarefa de construção, organização e representação do conjunto documental, sendo possível identificar o trabalho junto ao conjunto documental na etapa de armazenamento dos dados, na Figura 1 infracitada, sendo as etapas de coleta, descrição, análise e recoleta partes dessa atividade.

Figura 1 – O conjunto documental no ciclo de vida da pesquisa



Fonte: Universidade da Califórnia, Santa Cruz – Biblioteca Universitária⁵

Assim, a partir do exposto ao longo da pesquisa, foi investigado como os repositórios digitais integrados realizaram a tarefa de descrição dos seus recursos informacionais, relacionando com recursos digitais de outras fontes, constituindo, destarte, o conjunto documental apresentado nesta pesquisa.

⁵ Ciclo de vida da gestão dos dados de pesquisa: <https://guides.library.ucsc.edu/datamanagement>

4 PROCEDIMENTOS METODOLÓGICOS

Os procedimentos metodológicos, que refletiram a natureza do objeto de pesquisa e os objetivos propostos para esta investigação, serão apresentados nesta seção. Este estudo está focado em uma pesquisa descritiva acerca dos recursos para a organização e representação da informação para os ambientes digitais, mais especificamente, do esquema de metadados empregado para descrição do conjunto documental. Com o delineamento de uma pesquisa bibliográfica e documental, sendo esta última para analisar o ambiente digital selecionado, inclusive sua documentação e diretrizes para procedimentos de arquivamento e descrição dos metadados.

O estudo iniciou-se com uma pesquisa bibliográfica que buscou apresentar uma discussão sobre a organização e representação da informação em ambientes digitais, os recursos disponíveis para a descrição do conjunto documental, por meio de metadados, discorrer sobre a comunicação científica, incluindo o assunto de dados de pesquisa, dos repositórios institucionais, digitais, de dados de pesquisa e o repositório digital integrado.

A pesquisa bibliográfica é um levantamento que pode ser realizado em livros, artigos científicos, monografias, entre outros. E, de acordo com Gil (2009, p. 50), “[...] a principal vantagem da pesquisa bibliográfica reside no fato de permitir ao investigador a cobertura de uma gama de fenômenos muito mais ampla do que aquela que poderia pesquisar diretamente.”. Neste momento, a pesquisa buscou fornecer subsídios para a seleção e análise de publicações acerca dos repositórios institucionais, digitais, de dados de pesquisa e repositórios digitais integrados.

Nesta etapa, realizou-se uma pesquisa em bases de dados nacionais e internacionais por meio do uso de operadores booleanos e palavras-chave como repositório de dados de pesquisa, repositório digital, assim como em seus respectivos termos em inglês e espanhol - *research data repository*, *digital repository*, *repositorio de datos de investigación*, *repositorio digital*.

A seleção dos repositórios a serem estudados foi realizada a partir da análise de repositórios indexados no *Directory of Open Access Repositories* (OPENDOAR) e *Registry of Open Access Repositories* (ROAR). Estes buscadores indexam repositórios digitais (institucionais ou dados de pesquisa) em âmbito mundial, buscando identificar e comparar os manuais e diretrizes para o depósito dos dados de pesquisa e publicações e os recursos utilizados para a organização e representação

da informação do conjunto documental.

A seleção dos repositórios também teve um recorte geográfico, visando identificar e investigar os repositórios digitais integrados pertencentes às instituições de pesquisa dos países da América do Sul. Dentre a delimitação geográfica, os repositórios selecionados são aqueles que possuem tanto publicações científicas (como artigos, trabalhos em eventos, dissertação ou tese) quanto conjunto de dados de pesquisa.

Nesta etapa, realizada nos dias 17 e 18 de fevereiro de 2022, elaborou-se um formulário utilizando o *software* Excel para a coleta de dados, tais como: nome do repositório, endereço para acesso, país e o *software* utilizado. Primeiramente, para organizar e identificar os repositórios com as características de um repositório digital integrado e, também, para identificar os *softwares* mais utilizados.

Na segunda etapa, realizada entre os dias 1 e 3 de março de 2022, procedeu-se à exploração dos repositórios, a fim de identificar a existência de políticas e/ou diretrizes para o depósito dos diversos tipos de itens possíveis, assim como a tipologia empregada para a representação das publicações. Esta etapa forneceu uma visão mais geral sobre quais elementos cada repositório emprega, se são empregados mais de um esquema de metadados.

Na terceira etapa, entre os dias 3 e 4 de março de 2022, realizou-se a coleta das informações sobre o uso de dados abertos, a utilização de recursos para o relacionamento entre os elementos constituintes do conjunto documental, a utilização das práticas de dados conectados, o emprego de identificadores persistentes e a utilização de vocabulário controlado (para assuntos, tipologia dos arquivos). Nesta etapa, também foi empregado um formulário em Excel, com a nomeação dos repositórios, a identificação dos tipos de arquivos ou tipologias neles presentes, a investigação dos esquemas de metadados (explorando os princípios de modularidade e extensibilidade), o uso de vocabulário controlado em formato de dados conectados e a localização dos identificadores únicos para os diversos elementos, como autoria (ORCID) e identificação de recursos (DOI).

A quarta etapa, realizada no dia 7 de março de 2022, constituiu-se na coleta de informações sobre os campos de metadados empregados para a descrição do conjunto documental, buscando identificar os elementos presentes na descrição referente ao relacionamento entre os elementos do conjunto documental, tais como metadados de relacionamento, autoria e tipologia documental. A análise se voltou

para o uso de recursos e instrumentos que empregassem práticas para dados conectados e abertos, conforme apresentado por Ríos-Hilario, Martín-Campo, Ferreras-Fernández (2012), Isotani e Bittencourt (2015) e Sousa e Ramalho (2020). Assim, uma planilha Excel foi utilizada para coletar os campos de metadados empregados para cada relacionamento acima exposto.

A quinta etapa constituiu-se na análise das relações entre os dados coletados na terceira e quarta etapas. Dessa maneira, buscou-se identificar o uso de recursos e práticas que permitissem a relação entre o conjunto documental, autoria e tipologia, visando expor o uso de identificadores persistentes e vocabulários controlados para garantir maior consistência das informações utilizadas e contextualização da produção científica.

Para análise dos dados coletados, em todas as etapas, utilizou-se a Análise de Conteúdo (BARDIN, 1977), que fundamentou as etapas de seleção dos documentos e repositórios a serem analisados, por meio de uma pré-análise, seguida pela exploração dos conteúdos (incluindo localização de diretrizes para depósito, identificação de conjuntos de metadados e seus elementos e dos recursos digitais empregados) e, por fim, a coleta e a categorização dos dados.

5 ANÁLISE DOS DADOS E RESULTADOS

Os dados que constituem o corpus de análise foram coletados dos países que compõem a região da América do Sul e possuem repositórios registrados no OPENDOAR e no ROAR, sendo estes: Argentina, Bolívia, Brasil, Chile, Colômbia, Equador, Paraguai, Peru, Uruguai e Venezuela. Não existindo registro de repositórios nos países ou departamentos ultramarinos como os de Aruba, Curaçao, Guiana, Guiana Francesa e Suriname.

Assim, foi identificada a quantidade de 663 repositórios, divididos nos países, conforme a Tabela 1 abaixo.

Tabela 1 – Repositórios da América do Sul

PAÍS	Nº DE REPOSITÓRIOS
Argentina	73
Bolívia	3
Brasil	186
Chile	27
Colômbia	128
Equador	38
Paraguai	2
Peru	172
Uruguai	10
Venezuela	24

Fonte: OPENDOAR⁶ e ROAR⁷

Em seguida, foram selecionados os repositórios que possuíssem as características desejadas de um repositório digital integrado, isto é, que tivessem depósitos de trabalhos acadêmicos (monografias, dissertações e teses), publicações científicas (artigos científicos em periódicos científicos ou apresentações em eventos) e dados de pesquisa.

Após a coleta de dados, a análise foi separada em duas subseções. A primeira apresenta a estrutura dos repositórios, suas diretrizes de arquivamento, suas principais características, entre outras funções presentes. E a segunda contém a análise sobre os conjuntos de metadados utilizados e os recursos que representassem o conjunto documental.

⁶ OPENDOAR: <https://v2.sherpa.ac.uk/opendoar/>

⁷ ROAR: <http://roar.eprints.org/>

5.1 OS REPOSITÓRIOS DIGITAIS INTEGRADOS DA AMÉRICA DO SUL

Foram selecionados 32 repositórios digitais integrados, conforme as características citadas acima, resultando em repositórios da Argentina, Brasil, Chile, Colômbia, Peru e Uruguai, apresentados no Quadro 4 abaixo.

Quadro 4 – Repositórios da América do Sul

NOME DO REPOSITÓRIO	ENDEREÇO	PAÍS	SOFTWARE
Repositório Digital da Faculdade de Filosofia e Letras	http://repositorio.filo.uba.ar/	Argentina	DSPACE
Repositorio Institucional del CONICET	https://ri.conicet.gov.ar/	Argentina	DSPACE
INTA Digital	https://repositorio.inta.gob.ar/xmlui/	Argentina	DSPACE
Repositorio Institucional Digital de la Universidad Nacional de Río Negro	https://rid.unrn.edu.ar/	Argentina	DSPACE
Repositorio Digital de la FCEyS-UNMDP	http://nulan.mdp.edu.ar/	Argentina	Eprints
Producción Académica UCC	http://pa.bibdigital.ucc.edu.ar/	Argentina	Eprints
Repositorio Digital Institucional Universidad Nacional del Comahue	http://rdi.uncoma.edu.ar/	Argentina	DSPACE
Repositorio Digital da Universidade Nacional de Córdoba	https://rdu.unc.edu.ar/	Argentina	DSPACE
Repositorio Institucional de la UNLP	http://sedici.unlp.edu.ar/	Argentina	DSPACE
Umaza Digital	http://repositorio.umaza.edu.ar/	Argentina	DSPACE
Carpe dIEN	http://carpedien.ien.gov.br/	Brasil	DSPACE
Repositório Institucional da Fiocruz - ARCA	https://www.arca.fiocruz.br/	Brasil	DSPACE
Maxwell - PUC Rio	https://www.maxwell.vrac.puc-rio.br/	Brasil	Maxwell
Repositório Digital FGV	https://bibliotecadigital.fgv.br/dspace/	Brasil	DSPACE
Repositório Institucional UNESP	https://repositorio.unesp.br/	Brasil	DSPACE
Repositório Institucional da ENAP	https://repositorio.enap.gov.br/	Brasil	DSPACE
Repositório Institucional da UFAM	https://riu.ufam.edu.br/	Brasil	DSPACE
Repositório Institucional da Universidade de Brasília	https://repositorio.unb.br/	Brasil	DSPACE
Repositório do INPA	https://repositorio.inpa.gov.br/	Brasil	DSPACE
Repositório Institucional da Universidade Federal de São Carlos	https://repositorio.ufscar.br/	Brasil	DSPACE
Biblioteca Digital CIREN	http://bibliotecadigital.ciren.cl/	Chile	DSPACE
Repositorio UC	http://repositorio.uc.cl/	Chile	DSPACE
Repositorio Institucional Sêneca	https://repositorio.uniandes.edu.co/	Colômbia	DSPACE

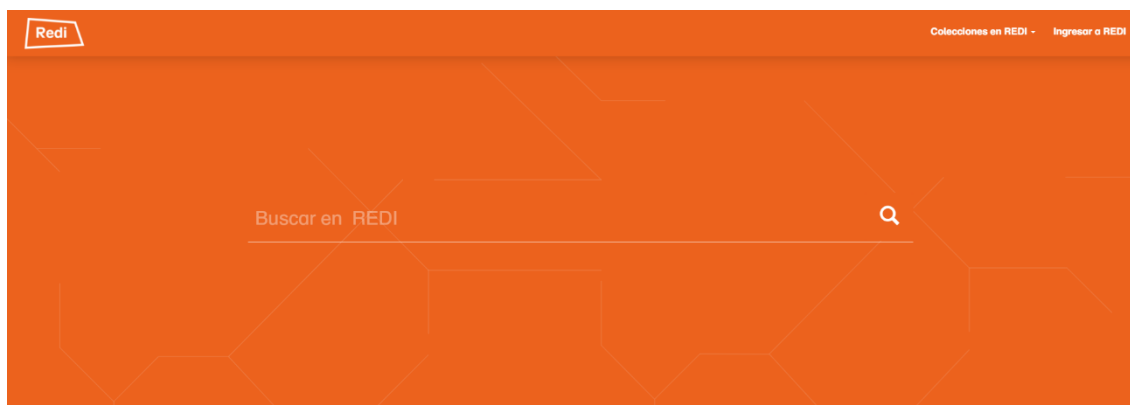
Repositorio Institucional Tecnológico de Antioquia	https://dspace.tdea.edu.co/	Colômbia	DSPACE
Repositorio Universidad Simón Bolívar	http://bonga.unisimon.edu.co/	Colômbia	DSPACE
Repositorio Universidad de La Costa	https://repositorio.cuc.edu.co/	Colômbia	DSPACE
Repositorio Universidad Autónoma de Manizales	https://repositorio.autonoma.edu.co/	Colômbia	DSPACE
Repositorio Académico UPC	https://repositorioacademico.upc.edu.pe/	Peru	DSPACE
Repositorio Institucional USIL	https://repositorio.usil.edu.pe/	Peru	DSPACE
Repositorio Institucional Universidade de Lima	http://repositorio.ulima.edu.pe/	Peru	DSPACE
Repositorio Institucional Universidad Peruana Cayetano Heredia	http://repositorio.upch.edu.pe	Peru	DSPACE
Repositorio Digital ANII	http://redi.anii.org.uy	Uruguai	DSPACE

Fonte: Dados da pesquisa

A primeira característica a ser percebida é a expressiva utilização do *software* livre DSPACE, dos 32 repositórios selecionados, 30 utilizam este *software*. Ivanovic e outros (2019) já haviam indicado uma grande preferência (42%) por este *software*, mas na América do Sul a utilização chega a 90%. Cabe ressaltar que o repositório da Pontifícia Universidade Católica do Rio de Janeiro (PUC-RIO) utiliza um sistema próprio, chamado de Sistema Maxwell, enquanto o outro sistema utilizado foi o Eprints.

A arquitetura informacional na qual os repositórios estão estruturados é muito parecida, mesmo aqueles repositórios que utilizam um *software* diferente do DSPACE, sendo que o único repositório com uma apresentação diferente dos demais é o *Repositorio Digital de la Agencia Nacional de Investigación e Innovación (ANII)*, do Uruguai, no qual não há organização por comunidades ou coleções e a página inicial é um campo de busca, conforme demonstra a Figura 2, a seguir.

Figura 2 – Página inicial do Repositório Digital ANII



Fonte: Repositório Digital ANII⁸

Esta etapa se constituiu, também, na coleta de algumas características dos repositórios, como a apresentação de diretrizes para o autoarquivamento, a organização da tipologia documental e o uso de esquemas de metadados, e se existiria a combinação no uso desses esquemas, como exposto no Quadro 5, infracitado.

Quadro 5 – Características dos repositórios da América do Sul

NOME DO REPOSITÓRIO	POSSUI DIRETRIZ PÚBLICA PARA AUTOARQUIVAMENTO?	ORGANIZA POR TIPO DE DOCUMENTO?	COMBINA ESQUEMAS DE METADADOS?	METADADO(S) EMPREGADO(S)
Repositório Digital da Faculdade de Filosofia e Letras	sim	sim	não	Dublin Core
Repositorio Institucional del CONICET	sim	sim	sim	Dublin Core; DataCite; Conicet
INTA Digital	não	sim	não	Dublin Core
Repositorio Institucional Digital de la Universidad Nacional de Río Negro	sim	sim	não	Dublin Core
Repositorio Digital de la FCEyS-UNMDP	sim	não	sim	Dublin Core; BIBO; FOAF
Producción Académica UCC	sim	sim	não	Dublin Core
Repositorio Digital Institucional Universidad Nacional del Comahue	sim	sim	não	Dublin Core
Repositorio Digital da Universidade Nacional de Córdoba	sim	sim	não	Dublin Core

⁸ Página inicial do Repositório digital ANII: <https://redi.anii.org.uy/jspui/>

Repositorio Institucional de la UNLP	sim	sim	sim	Dublin Core; Sedici; MODS
Umaza Digital	sim	sim	sim	Dublin Core; Umaza
Carpe dIEN	sim	sim	não	Dublin Core
Repositório Institucional da Fiocruz - ARCA	sim	sim	não	Dublin Core
Maxwell - PUC Rio	não	sim	não	Dublin Core
Repositório Digital FGV	sim	sim	não	Dublin Core
Repositório Institucional UNESP	sim	sim	não	Dublin Core
Repositório Institucional da ENAP	sim	sim	sim	Dublin Core; LOM
Repositório Institucional da UFAM	sim	sim	não	Dublin Core
Repositório Institucional da Universidade de Brasília	não	sim	não	Dublin Core
Repositório do INPA	sim	sim	não	Dublin Core
Repositório Institucional da Universidade Federal de São Carlos	sim	sim	não	Dublin Core
Biblioteca Digital CIREN	não	sim	não	Dublin Core
Repositório UC	sim	sim	não	Dublin Core
Repositório Institucional Sêneca	sim	sim	não	Dublin Core
Repositorio Institucional Tecnológico de Antioquia	não	sim	sim	Dublin Core; OAIRE
Repositorio Universidad Simón Bolívar	não	não	sim	Dublin Core; DataCite
Repositorio Universidad de La Costa	não	sim	não	Dublin Core
Repositorio Universidad Autónoma de Manizales	não	não	não	Dublin Core
Repositorio Académico UPC	sim	sim	sim	Dublin Core; RENATI

Repositorio Institucional USIL	não	sim	não	Dublin Core
Repositorio Institucional Universidade de Lima	sim	sim	sim	Dublin Core; RENATI
Repositorio Institucional Universidad Peruana Cayetano Heredia	sim	não	não	Dublin Core
Repositorio Digital ANII	não	sim	não	Dublin Core

Fonte: Dados da pesquisa

Conforme os dados no quadro acima, a sinalização verde corresponde aos repositórios que empregam as funções correspondentes na coluna, assim, do total de 32 repositórios da América do Sul analisados, 22 possuem em comum alguma apresentação pública de diretrizes ou tutoriais para o autoarquivamento; 28 repositórios permitem a organização do acervo em tipologias documentais; e todos os repositórios utilizam o Dublin Core como principal esquema de metadados.

A importância das diretrizes ou tutoriais serem públicos pode se basear na própria ideia de transparência e publicização nas práticas defendidas pela filosofia da Ciência Aberta. Dessa maneira, diferentes pesquisadores e profissionais de informação (como os bibliotecários ou os gerentes de repositórios) poderão compartilhar suas experiências, expor as suas práticas no povoamento do repositório, não ficando restritos à comunidade interna da instituição.

Uma questão relevante quanto à publicização dos Manuais se refere aos repositórios institucionais da UFSCAR, da Fiocruz – ARCA e da UNESP, pois possuem manuais específicos para suas tipologias documentais empregadas (Teses, dissertações, TCC, artigos, *dataset*, entre outros). Nesse sentido, destacam-se o arquivamento de Planos de Gestão de Dados (PGD) pela UNESP, o qual permite a identificação dos financiadores dos projetos, possibilitando, inclusive, a relação do PGD com o conjunto de dados gerados na pesquisa, e o autoarquivamento pelo repositório da Fiocruz – Arca, que pode ser realizado de forma manual, por uso de identificador (DOI, por exemplo) ou por um arquivo de referência (em XML).

Cabe ressaltar que os tutoriais/manuais do Repositório Institucional da Universidade Federal de São Carlos (RI-UFSCAR)⁹ e o da Fiocruz - ARCA¹⁰ indicam

⁹ Manual de autodepósito de dados de pesquisa RI-UFSCAR: <https://repositorio.ufscar.br/static/ri-ufscar-manual-autodeposito-dados.pdf>

¹⁰ Manual de tratamento de dados da Fiocruz – ARCA:

a possibilidade de se relacionar os dados de pesquisa depositados a uma publicação ou produção científica. O primeiro indica o item 23 em seu tutorial, enquanto que o segundo permite o relacionamento com mais de um objeto digital, indicando que este campo é um item repetitivo. Em ambos os casos, o relacionamento é realizado por uma URL ou URI.

O repositório institucional da *Universidad Nacional de La Plata*¹¹ possui uma página com informações que auxiliam os pesquisadores em diversas questões, como, por exemplo, adquirir um ORCID, citar documentos digitais e quais objetos digitais são possíveis de serem depositados. O *Repositorio Institucional del CONICET*¹², por sua vez, disponibiliza uma página com informações de ajuda sobre ciclo da publicação, métricas alternativas e sistemas de menção das publicações. Sobre comunicação científica e gestão de dados, o repositório da *Pontificia Universidad Católica de Chile*¹³ também disponibiliza uma página rica em informações para os pesquisadores, com o ciclo de vida da gestão de dados, um passo a passo de como gerir os dados, entre outras informações sobre pesquisa. O repositório da Fiocruz – Arca também se destaca ao disponibilizar uma página com diversas informações sobre a Ciência Aberta¹⁴, além de outras informações sobre dados abertos, acesso aberto, um glossário, entre outras informações.

A organização das tipologias documentais é um importante fator que auxilia na busca realizada no acervo do repositório, facilitando a localização do tipo de arquivo que o pesquisador tenha a necessidade, especialmente ao se tratar de um repositório com grande variedade de tipologias. Assim, quase 90% dos repositórios permitem que se localize uma tipologia específica, sendo que alguns, como o repositório digital da *Facultad de Ciencias Económicas y Sociales de la Universidad Nacional de Mar del Plata (FCEyS-UNMDP)*, permitem que na busca avançada se pesquise por tipologias e formatos de arquivos depositados, possibilitando, inclusive, combinar diversos formatos e refinar a pesquisa, conforme a Figura 3, infracitada.

https://www.arca.fiocruz.br/bitstream/iciict/13257/11/Manual_ARCA_versao_2019_FINALIZADO_Web.pdf

¹¹ Ajuda para os pesquisadores da UNLP: <http://sedici.unlp.edu.ar/pages/ayudaInvestigadores>

¹² Ajuda do repositório CONICET: <https://ri.conicet.gov.ar/wp/ayuda/>

¹³ Gestão de dados da UC: https://guiastematicas.bibliotecas.uc.cl/gdi_ip/inicio

¹⁴ Glossário Ciência Aberta: <https://portal.fiocruz.br/ciencia-aberta>

Figura 3 – Busca avançada Repositório FCEyS-UNMDP

Formato:	<input type="checkbox"/> HTML	<input type="checkbox"/> Video (WMV)	Tipo de Documento:	<input type="checkbox"/> Artículo	<input type="checkbox"/> Libro
	<input type="checkbox"/> PDF	<input type="checkbox"/> Video (MP4)		<input type="checkbox"/> Parte de Libro	<input type="checkbox"/> Tesis
	<input type="checkbox"/> Postscript	<input type="checkbox"/> Video (MP4)		<input type="checkbox"/> Monografía	<input type="checkbox"/> Preprint
	<input type="checkbox"/> Texto Plano	<input type="checkbox"/> Video (Flash)		<input type="checkbox"/> Documento de Conferencia	<input type="checkbox"/> Dataset
	<input type="checkbox"/> Rich Text (RTF)	<input type="checkbox"/> Video (AVCHD)			<input type="checkbox"/> Recurso de Aprendizaje
	<input type="checkbox"/> Microsoft PowerPoint	<input type="checkbox"/> XML			
	<input type="checkbox"/> Microsoft Excel	<input type="checkbox"/> N3			
	<input type="checkbox"/> Microsoft Word	<input type="checkbox"/> RDF/XML			
	<input type="checkbox"/> Imagen (JPEG)	<input type="checkbox"/> Archive (BZ2)			
	<input type="checkbox"/> Image (PNG)	<input type="checkbox"/> Archive (TGZ)			
	<input type="checkbox"/> Image (GIF)	<input type="checkbox"/> Archivo (ZIP)			
	<input type="checkbox"/> Image (BMP)	<input type="checkbox"/> Audio (WAV)			
	<input type="checkbox"/> Image (TIFF)	<input type="checkbox"/> Audio (MP3)			
	<input type="checkbox"/> Video (MPEG)	<input type="checkbox"/> Audio (OGG)			
	<input type="checkbox"/> Video (QuickTime)	<input type="checkbox"/> Audio (FLAC)			
	<input type="checkbox"/> Video (AVI)	<input type="checkbox"/> Audio (WMA)			
		<input type="checkbox"/> Otro			

Fonte: Repositório Digital da FCEyS-UNMDP¹⁵

A terceira característica analisada se refere ao uso de esquemas de metadados e à combinação entre diferentes formatos. Conforme Alves e Santos (2013), os metadados possuem os princípios da modularidade e da extensibilidade, permitindo que se construam conjuntos de metadados com elementos de diferentes esquemas e que se possa adicionar diferentes elementos aos conjuntos já existentes. Porém, conforme analisado, somente nove repositórios (menos de 30%) utilizam-se desses princípios.

Neste quesito, cabe ressaltar os repositórios da CONICET e da Universidad Simón Bolívar que combinam os esquemas de metadados: Dublin Core e DataCite. Estes dois esquemas de metadados, segundo diretrizes da OpenAIRE, são próprios para, respectivamente, repositórios institucionais e para dados de pesquisa. Portanto, os dois repositórios supracitados seguem as recomendações e implementaram o princípio da modularidade para combinar a representação do conjunto documental de suas produções científicas.

5.2 AS PRÁTICAS E OS RECURSOS PARA REPRESENTAÇÃO E ORGANIZAÇÃO DO CONJUNTO DOCUMENTAL

Nesta subseção, foram analisados os recursos e as boas práticas para a representação e a organização do conjunto documental, buscando identificar o emprego de práticas relacionadas aos dados abertos e conectados, à contextualização do conjunto documental, ao uso de identificadores persistentes e de

¹⁵ Busca avançada: <http://nulan.mdp.edu.ar/cgi/search/avanzada>

vocabulários controlados (inclusive para tipologia documental).

Assim, identificou-se quais repositórios indicavam a preferência pelo uso de dados abertos (especificando o tipo de formato do arquivo). Analisou-se, ainda, se explicitavam as relações entre as produções científicas, se utilizavam dados conectados para relacionar os documentos, quais identificadores persistentes eram empregados (autoria, relação do conjunto documental e tipologia documental) e se os repositórios utilizavam algum tipo de vocabulário controlado, em especial para determinar os termos próprios para as tipologia empregadas. Estas questões são analisadas a partir de subseções específicas no Quadro 6: dados abertos conectados, vocabulário controlado e conjunto documental.

Quadro 6 – Dados abertos conectados e outros recursos

NOME DO REPOSITÓRIO	USA DA?	USA DC?	ELO NO CD?	QUAIS IP?	UTILIZA VC?
Repositório Digital da Faculdade de Filosofia e Letras	sim	sim	sim	DOI; info:eu-repo; info:ar-repo	Tesouro BC; OpenAIRE; SNRD
Repositorio Institucional del CONICET	sim	sim	sim	DOI; info:eu-repo; info:ar-repo	OpenAIRE; SNRD
INTA Digital	sim	sim	sim	DOI; PURL; info:eu-repo; info:ar-repo	FORD; OpenAIRE; SNRD
Repositorio Institucional Digital de la Universidad Nacional de Río Negro	sim	sim	não	PURL	FORD
Repositorio Digital de la FCEyS-UNMDP	sim	sim	sim	PURL; info:eu-repo; info:ar-repo; ORCID	FORD; OpenAIRE; SNRD
Producción Académica UCC	sim	sim	sim	ORCID	não
Repositorio Digital Institucional Universidad Nacional del Comahue	sim	sim	sim	DOI; PURL	FORD
Repositorio Digital da Universidade Nacional de Córdoba	sim	sim	sim	DOI; PURL	FORD
Repositorio Institucional de la UNLP	sim	sim	sim	DOI	FORD
Umaza Digital	sim	sim	não	DOI	não
Carpe diEN	sim	sim	não	DOI; LATTES	não
Repositório Institucional da Fiocruz - ARCA	sim	sim	sim	DOI	DeCS; MeSH
Maxwell - PUC Rio	sim	sim	sim	DOI	não
Repositório Digital FGV	sim	não	não	Handle	Bibliodata
Repositório Institucional UNESP	sim	sim	sim	DOI; LATTES; ORCID; CAPES	Tesouro UNESP
Repositório Institucional da ENAP	sim	sim	sim	Handle; URL	ENAP
Repositório Institucional da UFAM	sim	sim	não	LATTES; ORCID	não

Repositório Institucional da Universidade de Brasília	sim	sim	sim	DOI; ORCID	não
Repositório do INPA	sim	sim	não	DOI	VCGE
Repositório Institucional da UFSCAR	sim	sim	sim	DOI; LATTES	CNPq
Biblioteca Digital CIREN	sim	sim	sim	Handle; URL	não
Repositório UC	sim	sim	sim	DOI	CDD
Repositório Institucional Sêneca	sim	sim	sim	PURL; info:eu-repo	COAR; DRIVER; REDCOL; LCSH
Repositorio Institucional Tecnológico de Antioquia	sim	sim	sim	PURL; info:eu-repo	COAR; DRIVER; REDCOL; MeSH; COAR; AGROVOC;
Repositorio Universidad Simón Bolívar	sim	sim	não	DOI; PURL; info:eu-repo	DataCite; DRIVER
Repositorio Universidad de La Costa	sim	sim	não	DOI	não
Repositorio Universidad Autónoma de Manizales	sim	sim	não	DOI; PURL; info:eu-repo	COAR; DRIVER; REDCOL
Repositorio Académico UPC	sim	sim	sim	DOI; PURL; ORCID	RENATI
Repositorio Institucional USIL	sim	sim	sim	DOI; PURL; info:eu-repo; info:pe-repo	RENATI; FORD; OCDE
Repositorio Institucional Universidade de Lima	sim	sim	sim	DOI; PURL; info:eu-repo; info:pe-repo	RENATI; FORD; OCDE
Repositorio Institucional Universidad Peruana Cayetano Heredia	sim	sim	sim	DOI; PURL; info:eu-repo; info:pe-repo	RENATI; FORD; OCDE
Repositorio Digital ANII	sim	sim	sim	DOI	não

LEGENDA: DA = Dados Abertos; CD = Conjunto Documental; DC = Dados Conectados; IP = Identificador Persistente; VC = Vocabulário Controlado.

Fonte: Dados da pesquisa

A sinalização verde representa os repositórios que empregam as boas práticas para dados abertos e conectados, além de empregarem *links* para relacionar os recursos digitais (dados de pesquisa, publicações científicas e produções acadêmicas) formando o conjunto documental. A sinalização amarela indica os repositórios que utilizam identificadores persistentes de vocabulários controlados para tipologia documental (ou tipos de recursos). Estas características serão analisadas nas subseções seguintes.

Destacam-se as diretrizes estabelecidas pelos órgãos governamentais que apresentam recomendações para a descrição e representação dos registros de metadados para os repositórios institucionais, sendo identificadas as da Argentina

(Diretrizes do Sistema Nacional de Repositórios Digitais, pelo SNRD¹⁶), do Brasil (Diretrizes OpenAIRE para repositórios de publicações científicas¹⁷ e dados¹⁸, pelo IBICT), da Colômbia (Diretrizes para repositórios institucionais¹⁹, pela Rede Colombiana) e do Peru (Diretrizes para repositórios institucionais²⁰, pelo CONCYTEC). Estas diretrizes irão auxiliar na análise dos campos de metadados empregados pelos repositórios, assim como os recursos por estes empregados.

5.2.1 Dados abertos conectados e identificadores persistentes

Segundo exposto por Isotani e Bittencourt (2015), os dados conectados não precisam ser essencialmente abertos, pois nem todos dados abertos estão conectados também. Dessa maneira, os autores apresentam a classificação das cinco estrelas para identificação das boas práticas para dados abertos e conectados, porém não é uma prática adotada pelos repositórios, mesmo que estes façam referências à importância do acesso aberto à produção científica.

Conforme o Quadro 6, supracitado, identificou-se os repositórios que utilizam, ao menos, uma boa prática básica pertencente à classificação de primeira estrela, isto é, os arquivos depositados não podem estar protegidos, permitindo o acesso livre, com exceção dos arquivos embargados²¹. Portanto, todos os repositórios disponibilizam (cor verde na primeira coluna) dados abertos para serem acessados.

A seguir, tem-se alguns exemplos:

- O manual para o autoarquivamento no repositório da *Facultad de Filosofía y Letras de la Universidad de Buenos Aires*²², o qual indica alguns formatos de arquivo, como: texto (PDF), dados (XML), imagem (TIFF, JPG e PNG), entre outros, porém sem explicitar a necessidade de arquivos abertos e não proprietários;
- O *Repositorio Institucional Digital de la Universidad Nacional de Río*

¹⁶ Diretriz do Sistema Nacional de Repositórios Digitais: https://repositoriosdigitales.mincyt.gob.ar/files/Directrices_SNRD_2015.pdf

¹⁷ Diretrizes OpenAIRE para Repositórios de publicações científicas: <https://livroaberto.ibict.br/handle/123456789/1087>

¹⁸ Diretrizes OpenAIRE para Repositórios de Dados: <https://livroaberto.ibict.br/handle/123456789/1086>

¹⁹ Diretriz para repositórios institucionais da rede colombiana: <https://redcol.minciencias.gov.co/vufind/Content/directrices>

²⁰ Diretriz para repositórios institucionais: <https://portal.concytec.gob.pe/index.php/guias-y-documentos-de-trabajo/item/235-guia-alicia>

²¹ Arquivos com restrição temporária a pedido do pesquisador

²² Procedimentos de autoarquivamento: <http://repositorio.filo.uba.ar/handle/filodigital/3265>

*Negro*²³, em sua política, já explicita a preferência por formatos de arquivos abertos e não proprietários, como: PDF, *Open Document Format* (ODF) e o ePUB;

- O Repositório Institucional da UFSCAR, em seu manual de autodepósito, recomenda o uso de arquivos em dados abertos (para dados de pesquisa), como XML, JSON, CSV, porém informa que aceita todos os formatos disponíveis.

Assim, percebe-se que os repositórios possuem políticas voltadas para o uso de formatos de acesso aberto, porém ainda permitem o arquivamento em formatos proprietários, como DOCX, XLSX, entre outros. Portanto, os repositórios atingem somente a classificação de duas estrelas do W3C (dados abertos e legíveis por máquina, mas ainda com uso de formato proprietário).

Já em relação ao uso de dados conectados (*Linked Data*), dos 32 repositórios analisados, somente o repositório FGV não explora este recurso, disponibilizando para os usuários somente o *handle* (URL) próprio do repositório como identificador persistente.

O uso de identificador persistente é um recurso importante para a relação entre conjunto de dados, publicações científicas, produções acadêmicas, entre outros objetos digitais, sendo os mais utilizados o DOI e o URI (quando o objeto digital não possui DOI ou quando a relação é feita com registros dentro do próprio repositório). Utilizou-se, também, o formato de URN, para indicar ISBN, ISSN, porém sem uso de dados conectados.

Outros identificadores persistentes utilizados são o de autoria, especificamente o ORCID e o LATTES. Estes, no entanto, foram utilizados em duas formas: em forma de caracteres literais e em forma de URI (HTTP). Não foi verificado o uso de outros identificadores de autoria, como o VIAF e o ResearcherID.

5.2.2 Vocabulário controlado

O vocabulário controlado é um registro de autoridade que permite determinar o termo preferido ao se descrever um recurso, sendo identificado por dois tipos de uso:

²³ Políticas del Repositorio Institucional Digital de la Universidad Nacional de Río Negro: <https://rid.unrn.edu.ar/rid-policies#formatos>

- **assunto** (em forma literal ou em forma do identificador persistente PURL);
- **tipologia documental** (em forma literal ou por uso de identificador persistente PURL, info:ar-repo, info:eu-repo, info:pe-repo).

Os vocabulários controlados que utilizam identificadores persistentes utilizam os recursos, conforme exposto por Fujita e Tolare (2019), por meio de campos de metadados relacionados a uma lista estruturada. Porém, em relação ao vocabulário de assuntos, somente um repositório utiliza os recursos e as práticas de dados conectados para tal, não sendo identificados exemplos citados por Ríos Hilario e Gil Urdiciain (2015) no uso de vocabulário controlado do Sistema de Classificação Decimal de Dewey por meio do esquema de metadados MADS.

Em relação aos vocabulários controlados para assuntos em lista estruturada, foram identificados três vocabulários controlados entre os repositórios, sendo o vocabulário da Biblioteca Central da FFyL, as Disciplinas Ford e o AGROVOC.

A Biblioteca Central da Faculdade de Filosofia e Letras disponibiliza um vocabulário eletrônico com os temas relacionados às produções acadêmicas do departamento e está organizado como demonstra a Figura 4 abaixo.

Figura 4 - Vocabulário da Biblioteca Central da FFyL

Vocabulario de la Biblioteca Central de la FFyL

Inicio Mi cuenta Buscar Búsqueda avanzada Sobre...

0-9 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

- ABORIGENES ▶
- ANTROPOLOGIA ▶
- ARTES ▶
- BIBLIOTECOLOGIA ▶
- CIENCIAS DE LA EDUCACION ▶
- CIENCIAS EXACTAS ▶
- CIENCIAS NATURALES ▶
- CIENCIAS SOCIALES ▶
- EDICION ▶
- FILOSOFIA ▶
- GEOGRAFIA ▶
- HISTORIA ▶
- INGENIERIA ▶
- LENGUA LATINA ▶
- LENGUAS BEREBERES ▶
- LENGUAS ROMANCES ▶
- LINGUISTICA ▶
- LITERATURA ▶
- MEDECINA ▶

Fonte: Vocabulario de la Biblioteca Central de la FFyL²⁴

Este vocabulário organiza os assuntos em temas mais gerais e, ao clicar no termo *Bibliotecologia*, abre-se uma lista com termos específicos da área, conforme descrito na Figura 5.






































²⁴ Vocabulário da BCFFyL: <https://www.vocabularyserver.com/filo/index.php>

Figura 5 – Termos específicos de *Bibliotecologia*

Términos alternativos

UP ↖ *Biblioteconomía*

Términos específicos

- TE1 ↓ [ACCESO A LA INFORMACION](#) 
- TE1 ↓ [ADQUISICION DE LIBROS](#) 
- TE1 ↓ [ANALISIS DOCUMENTAL](#)  ▶
- TE1 ↓ [ANALISIS FORMAL](#) 
- TE1 ↓ [ANUARIOS](#) 
- TE1 ↓ [AUTOMATIZACION DE BIBLIOTECAS](#) 
- TE1 ↓ [AUTORIDADES](#) 
- TE1 ↓ [BASE DE DATOS](#) 
- TE1 ↓ [BIBLIOFILIA](#) 
- TE1 ↓ [BIBLIOGRAFIAS](#)  ▶
- TE1 ↓ [BIBLIOMETRIA](#)  ▶
- TE1 ↓ [BIBLIOTECAS DIGITALES](#) 
- TE1 ↓ [BIBLIOTECAS VIRTUALES](#) 
- TE1 ↓ [BUSQUEDA BIBLIOGRAFICA](#)  ▶
- TE1 ↓ [CATALOGACION](#) 
- TE1 ↓ [CATALOGOS](#)  ▶
- TE1 ↓ [CONTROL BIBLIOGRAFICO](#) 
- TE1 ↓ [DESARROLLO DE LA COLECCION](#) 
- TE1 ↓ [DESCRIPCION BIBLIOGRAFICA](#) 
- TE1 ↓ [DISEMINACION DE LA INFORMACION](#) 
- TE1 ↓ [DOCUMENTACION](#)  ▶
- TE1 ↓ [ENCUADERNACION](#) 
- TE1 ↓ [EX LIBRIS](#) 
- TE1 ↓ [EXTENSION CULTURAL DE LAS BIBLIOTECAS](#) 
- TE1 ↓ [FORMACION DE USUARIOS](#)  ▶
- TE1 ↓ [FUENTES DE INFORMACION](#)  ▶
- TE1 ↓ [HISTORIA DE LA IMPRENTA](#)  ▶
- TE1 ↓ [HISTORIA DE LAS BIBLIOTECAS](#) 
- TE1 ↓ [HISTORIA DEL LIBRO](#)  ▶
- TE1 ↓ [LIBROS](#) 
- TE1 ↓ [LISTA DE ENCABEZAMIENTO DE MATERIAS](#) 
- TE1 ↓ [MANUSCRITOS](#)  ▶
- TE1 ↓ [MATERIALES NO LIBRARIOS](#)  ▶
- TE1 ↓ [MERCADEO](#) 
- TE1 ↓ [METADATOS](#) 
- TE1 ↓ [ORGANIZACION DE LA INFORMACION](#) 
- TE1 ↓ [POLITICA BIBLIOTECARIA](#)  ▶

Fonte: Vocabulário de la Biblioteca Central de la FFyL²⁵

Ao clicar em um termo específico, abre-se uma página com a indicação de

nota bibliográfica e os formatos de metadados disponíveis, entre estes, identificam-se os formatos DC e SKOS.

Figura 6 – Nota bibliográfica e metadados

VOCABULARIOS DE LENGUAJES DE INDIZACION

Inicio → BIBLIOTECOLOGIA → VOCABULARIOS DE LENGUAJES DE INDIZACION

Término Metadatos

VOCABULARIOS DE LENGUAJES DE INDIZACION

Nota bibliográfica ▼
 Tesouro de la Unesco, 1984

Términos genéricos

TG+ BIBLIOTECOLOGIA

VOCABULARIOS DE LENGUAJES DE INDIZACION

Inicio → BIBLIOTECOLOGIA → VOCABULARIOS DE LENGUAJES DE INDIZACION

Término Metadatos

Fecha de creación	02-Mar-2011
Término aceptado	02-Mar-2011
Términos descendent...	0
ARK	ark:/99152/t35yd8gxe1v0xo
Términos específicos	0
Términos alternativos	0
Términos relacionados	0
Notas	1
Metadatos	BS8723-5 DC MADS SKOS-Core VDEX XTM Zines JSON JSON-LD
Búsqueda	

Fonte: Vocabulário de la Biblioteca Central de la FFyL²⁶

O Ministério de Ciência, Tecnologia e Inovação, da Argentina, também disponibiliza um vocabulário controlado eletrônico relacionado às áreas de investigação e desenvolvimento, em inglês *Field of Research and Development* (FORD), baseado no manual da OCDE, de acordo com a Figura 7 infracitada.

Figura 7 – Classificação FORD



1 - Ciencias naturales

URI	https://purl.org/becyt/ford/1
Clasificación secundaria	<ul style="list-style-type: none"> 1.1 Matemáticas 1.2 Ciencias de la computación e información 1.3 Ciencias físicas 1.4 Ciencias químicas 1.5 Ciencias de la tierra y ciencias ambientales relacionadas 1.6 Ciencias biológicas 1.7 Otras ciencias naturales
Etiquetas	<ul style="list-style-type: none"> Ciencias naturales (es) Natural sciences (en) Ciências naturais (pt)

Fonte: Ministério de Ciência, Tecnologia e Inovação²⁷

A lista principal é formada pelas grandes áreas de: Ciências naturais, engenharia e tecnologia, ciências médicas e da saúde, ciências agrícolas e veterinárias, ciências sociais e humanidades e artes. Cada área tem um identificador persistente (PURL) expresso em uma URI e etiquetas com o termo em espanhol, inglês e português. Ao clicar em um termo específico, abre-se uma página com a mesma estrutura para este termo, como mostra a Figura 8:

²⁷ Classificação FORD: <https://biblioteca.mincyt.gob.ar/ford>

Figura 8 – Ciências de la computación e información



1.2 - Ciências de la computación e información

URI	https://purl.org/becyt/ford/1.2
Clasificación principal	1 Ciencias naturales
Etiquetas	Ciencias de la computación e información (es) Computer and information sciences (en) Ciências da computação e da informação (pt)
Incluye	Ciencias de la computación, ciencias de la información y bioinformática (desarrollo de hardware en 2.2, aspecto social en 5.8)
Conceptos y términos asociados (lista no exhaustiva)	administración de servidores; agrupamiento de datos; algoritmos; algoritmos genéticos; alineamiento de secuencias; alineamiento estructural de proteínas; almacén operacional de datos; almacenamiento de la información; análisis de datos; animaciones por computadoras; anotación de genomas; aprendizaje automático; árboles de decisión; bases de datos; bigdata; biocomputación; bioinformática; biología computacional; biología de sistemas; biología evolutiva computacional; búsqueda y recuperación de información; ciberdefensa; ciberseguridad; ciencias de la computación; ciencias de la computación aplicadas; ciencias de la información (aspectos informáticos); ciencias de la Información y bioinformática; ciencias de la informática; ciencias informáticas; cluster de datos; computación; computación aplicada a datos biológicos; computación concurrente; computación distribuida; cómputo científico; condificación binaria; conjunto de datos; conjuntos de datos; correo electrónico; datapapers; dataset; datos biológicos; datos genéticos; datos genómicos; datos geográficos; datos georreferenciados; datos satelitales; diseño asistido por computadora; diseño de computadoras; diseño gráfico por computadoras; encriptación; exploración de datos; georreferenciación de datos; gestión de datos; gestión de datos biológicos; grafos; industria de videojuegos; informática; informática biológica; informática en salud; ingeniería de software; interacciones proteína-proteína; Internet; Intranet; KDD; lenguajes de programación; lógica booleana; lógica computacional; medición de la biodiversidad; metadatos; minería de datos; minería de datos espaciales; modelo computacional; montaje genómico; multiprocesamiento; predicción de estructuras proteicas; predicción de genes; predicción de la expresión génica; predicción genética; procesamiento de datos; procesamiento de imágenes; procesamiento de la información; procesamiento de lenguaje natural; quimiometría; redes de computadoras; redes neuronales; reglas de asociación de datos; regresión lineal; rendimiento computacional; reconocimiento de patrones; representación de la información; representación del conocimiento; seguridad de la información; seguridad informática; simulación de sistemas biológicos; sistemas computacionales; sistemas de gestión de bases de datos; sistemas expertos; sistemas inteligentes; sistemas operativos; software; software astronómico; software libre; técnicas de minería de datos; tecnología de datos biológicos; teoría de la complejidad computacional; teoría de la computación; teoría de la información y códigos; visión artificial;

Fonte: Ministério de Ciência, Tecnologia e Inovação²⁸

Além da estrutura com URI e etiquetas, estão presentes conceitos e termos associados em uma lista exhaustiva.

A Organização das Nações Unidas para Alimentação e Agricultura disponibiliza um tesouro eletrônico multilíngue, chamado de AGROVOC, organizado em 25 assuntos principais: atividades, entidades, eventos, fatores, características, grupos, localização, medidas, métodos, objetos, fenômeno, processos, produtos, propriedades, recursos, local, estágio, estado, estratégias, assuntos, substâncias, sistemas, tecnologia e tempo. Cada um desses assuntos está subdividido em diversos assuntos mais específicos. Todos os assuntos possuem um termo preferido, conceitos mais específicos, nota de escopo, traduções e uma URI, conforme explicitado na Figura 9.

²⁸ Classificação FORD – classe específica: <https://biblioteca.mincyt.gov.ar/ford/1.2>

Figura 9 – Tesouro AGROVOC

The screenshot shows the AGROVOC thesaurus interface. On the left, there is a hierarchical tree with 'subjects' selected. The main area displays the following information for the 'subjects' concept:

- PREFERRED TERM:** subjects
- NARROWER CONCEPTS:** humanities (en), sciences (en), statistics as science (en)
- SCOPE NOTE:** Disciplines of study (en)
- IN OTHER LANGUAGES:**
 - ① 主题 (Chinese)
 - ① subjekty (Czech)
 - ① sujet (French)
 - ① Fachgebiet (German)
 - ① Fachrichtung (German)
 - ① Argomenti (Italian)
 - ① Subjek (Malay)
 - ① fag (Norwegian Bokmål)
 - ① области знания (Russian)
 - ① области учења (Serbian)
 - ① предмети (Serbian)
 - ① odborné oblasti (Slovak)
 - ① Sujetos (Spanish)
 - ① konu (Turkish)
 - ① дисципліни (Ukrainian)
- URI:** http://aims.fao.org/aos/agrovoc/c_330829
- Download this concept:** RDF/XML TURTLE JSON-LD (Created 9/25/08, last modified 11/24/21)
- EXACTLY MATCHING CONCEPTS:** <http://www.eionet.europa.eu/gemet/concept/14848>

Fonte: AGROVOC²⁹

Esse vocabulário também pode ser acessado em sua forma de metadados (RDF), onde pode-se identificar o uso dos esquemas de metadados SKOS e DC, como demonstra a Figura 10 abaixo.

Figura 10 – AGROVOC metadados

http://aims.fao.org/aos/agrovoc/c_330829	
subjects	
Property	Value
rdf:type	skos:Concept
skos:inScheme	http://aims.fao.org/aos/agrovoc/conceptScheme_7a97495f
skos:topConceptOf	http://aims.fao.org/aos/agrovoc
skos:narrower	http://aims.fao.org/aos/agrovoc/c_7379 http://aims.fao.org/aos/agrovoc/c_330830 http://aims.fao.org/aos/agrovoc/c_37989
skos:exactMatch	http://www.eionet.europa.eu/gemet/concept/14848
Property	Value
dcterms:created	2008-09-25T00:00:00Z
dcterms:modified	2021-11-24T18:10:26
void:inDataset	http://aims.fao.org/aos/agrovoc/void.ttl#Agrovoc
skos:scopeNote	Disciplines of study

Fonte: AGROVOC³⁰

Além dos vocabulários controlados para assunto, foram identificados os vocabulários para a tipologia documental (ou tipo de recursos) empregados nos repositórios. Cabe destacar os repositórios que utilizam identificadores persistentes para relacionar a tipologia documental a um vocabulário controlado em uma lista estruturada em forma de metadados. Os recursos identificados nos repositórios foram

²⁹ Tesouro AGROVOC: <https://agrovoc.fao.org/browse/agrovoc/en/>

³⁰ AGROVOC metadados: http://aims.fao.org/aos/agrovoc/c_330829.html

o uso de PURL, info:ar-repo, info:eu-repo e info:pe-repo.

Analisando o uso de PURL para classificar a tipologia documental empregada, percebe-se que apenas seis repositórios seguem as diretrizes de seus respectivos órgãos sobre a gestão da informação nos repositórios institucionais. Sendo três repositórios da Colômbia (*Repositorio Sêneca*, *Repositorio Institucional Tecnológico de Antioquia* e *Repositorio Universidad Autónoma de Manizales*) e três do Peru (*Repositorio Académico UPC*, *Repositorio Institucional USIL* e *Repositorio Institucional Universidad Peruana Cayetano Heredia*), estando representados no Quadro 7 abaixo.

Quadro 7 – PURL da tipologia documental / tipo de recurso

Nome do repositório	PURL	Metadado	Tipologia	Vocabulário	Diretriz
Repositório Institucional Sêneca	http://purl.org/coar/resource_type/c_2d8fbb1	dc.type.coar	Artigo científico	COAR	4.11 - REDCOL
	http://purl.org/redcol/resource_type/ART	dc.type.redcol	Artigo científico	REDCOL	4.11 - REDCOL
Repositorio Institucional Tecnológico de Antioquia	http://purl.org/coar/resource_type/c_dbb1	dc.type.coar	Conjunto de dados	COAR	4.11 - REDCOL
Repositorio Universidad Autónoma de Manizales	http://purl.org/coar/resource_type/c_bdcc	dc.type.coar	Tese	COAR	4.11 - REDCOL
	https://purl.org/redcol/resource_type/TM	dc.type.redcol	Tese	REDCOL	4.11 - REDCOL
Repositorio Académico UPC	http://purl.org/pe-repo/renati/type#tesis	renati.type	Tese	RENATI	8.2.30 - ALICIA
Repositorio Institucional USIL	https://purl.org/pe-repo/renati/type#trabajoDeInvestigacion	renati.type	Dissertação	RENATI	8.2.30 - ALICIA
Repositorio Institucional UPCH	http://purl.org/pe-repo/renati/type#trabajoDeSuficienciaProfesional	renati.type	Monografia	RENATI	8.2.30 - ALICIA

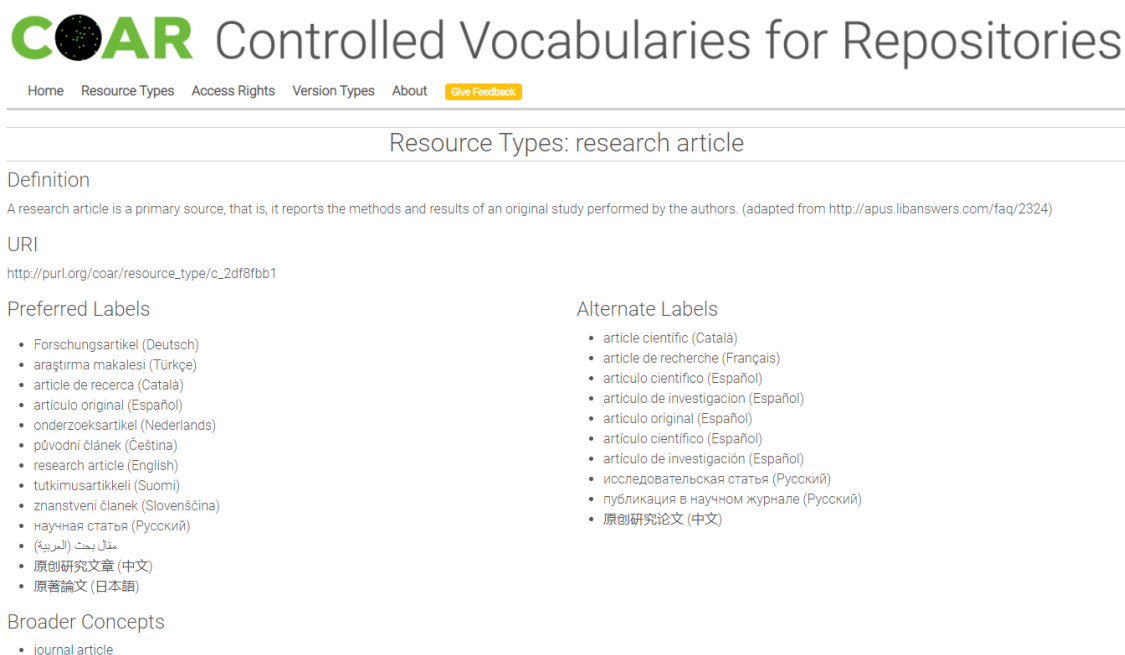
Fonte: Dados da pesquisa

No quadro supracitado encontram-se o nome do repositório, a PURL utilizada para identificar a tipologia em um vocabulário, o campo de metadado empregado, o nome do recurso conforme o vocabulário, o nome do vocabulário e a diretriz na qual foi baseada esta descrição.

A diretriz da Rede Colombiana de Informação Científica indica duas formas de se representar o recurso por meio de uma PURL, pode-se utilizar o vocabulário próprio da rede ou o vocabulário do COAR, devendo-se identificar no campo de metadados o vocabulário utilizado, sendo, respectivamente, “dc.type.redcol” ou “dc.type.coar”.

Ao acessar a PURL do COAR, tem-se acesso à definição da tipologia, à URI empregada como identificador persistente, às etiquetas preferidas e alternativas e, quando existir, à tipologia geral à qual esta pertença, como mostra a Figura 11, a seguir.

Figura 11 – Tipologia para recursos COAR



COAR Controlled Vocabularies for Repositories

Home Resource Types Access Rights Version Types About Give Feedback

Resource Types: research article

Definition

A research article is a primary source, that is, it reports the methods and results of an original study performed by the authors. (adapted from <http://apus.libanswers.com/faq/2324>)

URI

http://purl.org/coar/resource_type/c_2df8fbb1

Preferred Labels

- Forschungsartikel (Deutsch)
- araştırma makalesi (Türkçe)
- article de recerca (Català)
- artículo original (Español)
- onderzoeksartikel (Nederlands)
- původní článek (Čeština)
- research article (English)
- tutkimusartikkeli (Suomi)
- znanstveni članek (Slovenščina)
- научная статья (Русский)
- مقال بحث (العربية)
- 原創研究文章 (中文)
- 原著論文 (日本語)

Alternate Labels

- article científico (Català)
- article de recherche (Français)
- artículo científico (Español)
- artículo de investigación (Español)
- artículo original (Español)
- artículo científico (Español)
- artículo de investigación (Español)
- исследовательская статья (Русский)
- публикация в научном журнале (Русский)
- 原創研究论文 (中文)

Broader Concepts

- [journal article](#)

Fonte: Vocabulário controlado para repositórios do COAR³¹

Em relação ao vocabulário controlado próprio da Rede Colombiana de Informação Científica, este não foi encontrado ao acessar a URI, dando acesso ao repositório da rede, também não sendo possível encontrar a referência para este vocabulário. Entretanto, a diretriz da Rede Colombiana indica, na seção 4.11, a utilização dos dois vocabulários supracitados e os termos equivalentes entre si.

Já o Vocabulário RENATI³² para tipologia empregado pelos repositórios peruanos é de construção própria, não indicando equivalências entre outros vocabulários, além de possuir uma quantidade limitada de tipologias, sendo: Tese, *Trabajo de investigación* (dissertação), *Trabajo de suficiencia profesional* (monografia) e *Trabajo académico*, que pode ser uma variedade de tipologias de produção acadêmica (projetos, ensaios, investigação, entre outros).

Outra forma de controle de vocabulário para as tipologias e tipos de recursos também se encontra nas diretrizes já apresentadas, porém, somente é empregada nos repositórios da Argentina, Colômbia e Peru.

As três diretrizes explicitam que seguem as regras do COAR, assim como a diretriz do IBICT, e recomendam o uso de identificadores persistentes próprios. Dessa

³¹ Vocabulário Controlado do COAR: https://vocabularies.coar-repositories.org/resource_types/c_2df8fbb1/

³² Vocabulário controlado para tipos de trabalhos de investigação (RENATI): https://catalogos.concytec.gob.pe/vocabulario/renati_type.html

maneira, os recursos são:

- Argentina: info:ar-repo/semantics/ (SNRD) e info:eu-repo/semantics/ (COAR);
- Colômbia: info:eu-repo/semantics/ (COAR/DRIVER);
- Peru: info:eu-repo/semantics/ e info:pe-repo/semantics (COAR/DRIVER).

No Quadro 8, infracitado, são apresentadas as formas como os repositórios representaram as suas tipologias, sendo selecionados alguns exemplos para a demonstração.

Quadro 8 – Tipologias info:ar-repo, info:eu-repo e info:pe-repo

Nome do repositório	Vocabulário info:	Metadado	Tipologia	Vocabulário	Diretriz
Repositório Digital da FFyL	info:eu-repo/semantics/conferenceObject	dc.type	Artigo de evento	COAR	SNRD
	info:ar-repo/semantics/documento de conferencia	dc.type	Artigo de evento	SNRD	SNRD
Repositorio Institucional del CONICET	info:eu-repo/semantics/bookPart	dc.type	Capítulo de livro	COAR	SNRD
	info:ar-repo/semantics/parte de libro	dc.type	Capítulo de livro	SNRD	SNRD
INTA Digital	info:eu-repo/semantics/other	dc.type	Conjunto de dados	COAR	SNRD
	info:ar-repo/semantics/conjunto de datos	dc.type	Conjunto de dados	SNRD	SNRD
Repositorio Digital de la FCEyS-UNMDP	info:eu-repo/semantics/bachelorThesis	dc.type	Monografia	COAR	SNRD
	info:ar-repo/semantics/tesis de grado	dc.type	Monografia	SNRD	SNRD
Repositório Institucional Sêneca	info:eu-repo/semantics/bachelorThesis	dc.type.driver	Monografia	DRIVER	REDCOL
Repositorio Institucional Tecnológico de Antioquia	info:eu-repo/semantics/masterThesis	dc.type.driver	Dissertação	DRIVER	REDCOL
Repositorio Universidad Simón Bolívar	info:eu-repo/semantics/book	dc.type.driver	Livro	DRIVER	REDCOL
Repositorio Universidad Autónoma de Manizales	info:eu-repo/semantics/doctoralThesis	dc.type	Tese	COAR	REDCOL
	info:eu-repo/semantics/doctoralThesis	dc.type.driver	Tese	DRIVER	REDCOL
Repositorio Institucional USIL	info:eu-repo/semantics/book	dc.type	Livro	COAR	ALICIA
Repositorio Institucional Universidade de Lima	info:eu-repo/semantics/bachelorThesis	dc.type	Monografia	COAR	ALICIA
Repositorio Institucional UPCH	info:eu-repo/semantics/conferenceObject	dc.type	Artigo de evento	COAR	ALICIA
	info:pe-repo/semantics/software	dc.type	Software	COAR	ALICIA

Fonte: Dados da pesquisa

Percebe-se, portanto, que os repositórios apresentados no quadro supracitado seguem as recomendações relacionadas ao controle de vocabulários para as tipologias empregadas, existindo somente a divergência quanto aos metadados empregados pelos repositórios colombianos, que utilizam o campo “dc.type.driver”, o qual, segundo a própria diretriz da Rede Colombiana de Informação Científica, encontra-se obsoleto.

5.2.3 Conjunto documental

Ao se pensar em um conjunto documental, deve-se levar em consideração a relação entre os diversos elementos que constituem a representação dos objetos digitais e as relações existentes e possíveis com objetos digitais internos ou externos ao repositório. Para isso, além dos recursos e prática para o uso de dados abertos e conectados para a identificação de autores, assuntos e tipologias, deve-se investigar as relações que os diferentes tipos de recursos possuem.

Todas as diretrizes apresentadas anteriormente foram analisadas e identificados os recursos disponíveis e as práticas recomendadas. Dessa maneira, primeiramente, buscou-se identificar quais repositórios relacionam os seus objetos digitais com outros tipos de recursos, constatando-se que dos 32 repositórios, somente nove não apresentaram esse tipo de relação, não utilizando *links* (URL ou URI) entre os recursos.

Cabe ressaltar que, mesmo com 31 repositórios os quais utilizaram algum tipo de dados conectados, a falta de um tutorial próprio, ou mesmo a apresentação das recomendações das diretrizes anteriormente apresentadas, dificulta o uso mais generalizado destes procedimentos e de recursos percorridos por Ríos Hilario e Gil Urdiciain (2015), Lima (2019), entre outros, para se relacionar os dados, produções acadêmicas e publicações científicas com os registros de metadados dos recursos. Como exemplos, tem-se os seguintes casos:

- O registro do conjunto de dados³³, no repositório da UNRN, possui o campo “dc.relation.relatedProject”, que poderia utilizar uma URI para conectar os dados de pesquisa coletados ao projeto relacionado e depositado no sistema, porém utiliza caracteres literais para descrever a relação entre os dois documentos;
- O registro do conjunto de dados³⁴, no Repositório Institucional Sêneca, possui o campo “dc.relation.ispartof”, utilizando caracteres literais para descrever a relação com um artigo, mas poderia utilizar uma URI para conectar as duas produções científicas;
- Outro exemplo, no Repositório Institucional da ENAP, o registro do conjunto de dados³⁵ possui o campo “dc.description.abstract”, onde se faz a relação a uma publicação e a um serviço de forma literal, ao invés de se usar uma URI.
- Como último exemplo, um trabalho apresentado em um evento, depositado no Repositório da UPCH³⁶, utiliza o campo

³³ Registro dos metadados no repositório da UNRN: <https://rid.unrn.edu.ar/handle/20.500.12049/2532?mode=full>

³⁴ Registro dos metadados no repositório Sêneca: <https://repositorio.uniandes.edu.co/handle/1992/45861?show=full>

³⁵ Registro dos metadados no repositório da ENAP: <https://repositorio.enap.gov.br/handle/1/3695?mode=full>

³⁶ Registro dos metadados no repositório da UPCH: <https://repositorio.upch.edu.pe/handle/20.500.12866/10153?show=full>

“dc.relation.conference” sem uma URI para relacionar o registro ao evento.

Portanto, a organização e a representação da informação de um conjunto documental dependem da correta utilização de recursos digitais e dos campos de metadados e, para isso, é imprescindível a participação dos administradores dos repositórios e dos profissionais de informação por meio da elaboração e divulgação de tutoriais e manuais para o autodepósito. Esta questão recai sobre a publicização e o compartilhamento de experiências já existentes e, também, a adequação às diretrizes das agências governamentais que prestam serviços de informação e tecnologia.

Todas as diretrizes supracitadas apresentam recomendações para os campos a serem preenchidos, com indicação para preenchimento obrigatório, recomendado, opcional nos diversos campos que possibilitam a construção e a representação do conjunto documental.

Tendo estas questões em perspectiva, o Quadro 9, abaixo, apresenta os diversos campos de metadados que foram identificados realizando algum tipo de relação com outros recursos ou entidades, utilizando ou não dados conectados (URI, URL, DOI, PURL, entre outros). A intenção inicial foi de poder observar de forma mais geral como os repositórios identificavam as relações entre os diversos recursos digitais e de que forma essas relações eram expressas.

Quadro 9 – Campos de metadados e o conjunto documental

Função	Campo metadado	Função	Campo metadado	Função	Campo metadado
R E L A C I O N A M E N T A L C O N J U N T O	dc.description	A U T O R I A D O T R A B A L H O	dc.contributor	T I P O L O G I A	dc.datacite.DescriptionType
	dc.description.abstract		dc.contributor.advisor		dc.datacite.ResourceTypeGeneral
	dc.description.origen		dc.contributor.advisor1		dc.subtype
	dc.description2		dc.contributor.advisor1Lattes		dc.type ²
	dc.identifier		dc.contributor.advisorco		dc.type.coar
	dc.identifier.anii		dc.contributor.author ²		dc.type.content
	dc.identifier.citation		dc.contributor.authorlattes		dc.type.driver
	dc.identifier.doi		dc.contributor.coordinator		dc.type.other
	dc.identifier.other		dc.contributor.member		dc.type.redcol
	dc.identifier.scopus		dc.contributor.other		dc.type.version
	dc.identifier.uri		dc.creator		renati.type
	dc.relation		dc.creator.affiliation		sedici.subtype
	dc.relation.alternativeid		dc.datacite.affiliation		dc.datacite.ContributorType
	dc.relation.conference		dc.datacite.ContributorName		
	dc.relation.haspart		dc.datacite.Creator		
	dc.relation.hasversion		dc.description		
	dc.relation.isbasedon		dc.description.affiliation		
	dc.relation.isformatof		dc.description.fil		
	dc.relation.ispartof		dc.identifier.lattes		
	dc.relation.ispartofjournal		dc.identifier.orcid		
	dc.relation.publisherversion		dc.rights.holder		
	dc.relation.references		renati.advisor.dni		
	dc.relation.relatedProject		renati.advisor.orcid		
	dc.relation.uri		renati.author.dni		
	dc.relation.url		renati.juror		
	dc.relationtype.isSourceOf		sedici.creator.person		
dc.rights.license	unesp.author.orcid				
dc.source					
sedici.identifier.other					
sedici.identifier.uri					
sedici.relation.isRelatedWith					

Fonte: Dados da pesquisa

Primeiramente, percebe-se que os repositórios usam grande variedade de campos de metadados, especialmente para a função que relaciona o conjunto documental, porém, é possível identificar campos de metadados que não estão presentes nas diretrizes apresentadas, como exemplo tem-se: “dc.description.origen”, “dc.description2”, “dc.relation.conference”, “dc.isbasedon”, “dc.relation.ispartofjournal”, entre outros. Além disso, há variedade de formas para autoria e algumas variações para a tipologia.

Em segundo lugar, a recomendação para o preenchimento dos campos de relações entre os recursos é o uso de URI, isto é, a utilização de dados conectados que sejam legíveis por máquinas, porém, esta recomendação nem sempre é seguida. Assim, todos os campos do quadro supracitado foram investigados e selecionados somente aqueles que seguiram as recomendações de preenchimento. Além dos campos para relacionamento entre os dados de pesquisa, publicações científicas e produções acadêmicas, selecionou-se também os outros campos que utilizassem dados conectados por URI, URL, PURL ou outro identificador persistente tanto para

autoria quanto para tipologia, resultando no Quadro 10 infracitado.

Quadro 10 – Campos que utilizam dados conectados

Função	Campo de metadados	IP	Função	Campo de metadados	IP	Função	Campos de metadados	IP
Conjunto Documental	dc.identifier.doi	DOI	Autoria	dc.contributor.advisor1Lattes	LATTES	Tipologia	dc.type ²	PURL; ar-repo; eu-repo; pe-repo
Conjunto Documental	dc.identifier.other	URI	Autoria	dc.contributor.authorlattes	LATTES	Tipologia	dc.type.coar	PURL; eu-repo
Conjunto Documental	dc.identifier.uri	URI	Autoria	dc.identifier.lattes	LATTES	Tipologia	dc.type.driver	eu-repo
Conjunto Documental	dc.relation	URI	Autoria	dc.identifier.orcid	ORCID	Tipologia	dc.type.redcol	PURL
Conjunto Documental	dc.relation.alternativeid	URI	Autoria	renati.advisor.dni	RENATI	Tipologia	dc.type.version	PURL
Conjunto Documental	dc.relation.haspart	URI	Autoria	renati.advisor.orcid	ORCID	Tipologia	renati.type	PURL; pe-repo
Conjunto Documental	dc.relation.hasversion	URI	Autoria	unesp.author.orcid	ORCID			
Conjunto Documental	dc.relation.isbasedon	URI						
Conjunto Documental	dc.relation.ispartof	URI						
Conjunto Documental	dc.relation.relatedProject	URI						
Conjunto Documental	dc.relation.uri	URI						
Conjunto Documental	dc.relation.url	URI						
Conjunto Documental	dc.relationtype.isSourceOf	URI						
Conjunto Documental	sedici.identifier.other	URI						
Conjunto Documental	sedici.identifier.uri	URI						
Conjunto Documental	sedici.relation.isRelatedWith	URI						

Fonte: dados de pesquisa

Assim, dos 31 campos de metadados utilizados para relacionar o conjunto documental, 16 utilizam dados conectados no formato URI, URL ou DOI. Na função para autoria, o identificador Lattes e ORCID foram utilizados no formato HTTP (URI) e função de tipologia, somente um não utilizou um identificador PURL (em formato HTTP).

A partir da análise dos campos de metadados empregados, percebe-se a importância da noção de publicação ampliada, apresentada por Sales, Souza e Sayão (2014), para a representação do conjunto documental, permitindo, inclusive, que o usuário localize e identifique as publicações relacionadas em consonância com um dos princípios FAIR, o de encontrabilidade, conforme exposto por Ribeiro (2019), ou seja, identificadores persistentes para os dados, documentos e metadados. É importante ressaltar que este último auxilia tanto os usuários quanto a leitura por máquina e indexação em uma base de dados que permita a sua busca.

Corroborar-se, também, as afirmações de Ríos-Hilario, Martín-Campo, Ferreras-Fernández (2012) e Lima (2020) de que a estruturação de uma *web* semântica por meio de dados conectados auxilia na interoperabilidade entre sistemas de informação e o compartilhamento de objetos digitais para as pesquisas, sendo

importante a publicização, por parte dos repositórios, via tutoriais ou manuais, das experiências existentes para tal.

Porém, cabe ressaltar que a documentação para os dados de pesquisa, um dos recursos apresentados por diversos pesquisadores, como Pampel e outros (2013) e Crosas (2015), ainda é pouco utilizada, sendo indicada somente quando se discute a produção de um plano de gestão de dados por alguns repositórios.

6 CONSIDERAÇÕES FINAIS

Nesta pesquisa, buscou-se apresentar os resultados de uma investigação sobre a representação e a organização do conjunto documental relacionado às publicações e produções científicas. Para isso, recorreu-se às pesquisas na literatura científica da área da Ciência da Informação, especialmente nas subáreas de Organização e Representação da Informação e Comunicação Científica.

Para isso, tomou-se como base os trabalhos clássicos de Bräscher e Café (2008), Café e Sales (2010) e Maimone, Silveira e Tálamo (2011), por exemplo, a fim de dar suporte à discussão acerca da organização da informação, além da utilização de investigações e publicações mais recentes acerca da organização da informação em âmbito digital, como as pesquisas de Fujita e Tolare (2019), Lima (2020), Ríos Hilario e Gil Urdiciain (2015) e Sousa e Ramalho (2020).

Cortês (2006), Garvey (1979) e Meadows (1999) auxiliaram na discussão sobre a comunicação científica, sua transformação ao longo da história, embasando as práticas contemporâneas para divulgação e disseminação da produção científica, especialmente com foco nas pesquisas, publicações e dados abertos, conforme exposto por Aventurier e Alencar (2016), Callaghan (2019), Ribeiro (2019) e Sayão e Sales (2016a, 2016b).

Todos os autores e pesquisadores foram essenciais para se discutir e fundamentar as análises que reuniram dois campos do conhecimento (Organização e Representação da Informação e Comunicação Científica) os quais, cada vez mais, se tornam próximos e auxiliam bibliotecários, arquivistas e demais profissionais de informação a descreverem e disponibilizarem a informação científica para um grupo maior de pesquisadores.

Percebeu-se que os repositórios digitais têm recorrido a uma variedade de estratégias e práticas para o desenvolvimento de um ambiente digital capaz de armazenar diversos objetos digitais, com a percepção da necessidade de se seguir a filosofia do movimento de acesso aberto, que se iniciou com o acesso livre às publicações, culminando no acesso livre aos dados de pesquisa, às metodologias empregadas na pesquisa, aos *software* livres (tanto para coleta quanto para a análise dos dados), ao trabalho cooperativo entre pesquisadores e à devida identificação das autorias.

Os órgãos governamentais responsáveis por diretrizes, fomento e discussões

acerca da informação científica, o seu devido tratamento, entre outras iniciativas, têm uma responsabilidade crucial para que os repositórios empreguem corretamente os esquemas de metadados para descrever e representar o conjunto documental estudado nesta investigação, porém, percebe-se que não há um uso padronizado dos campos de metadados, podendo acarretar confusões por parte dos usuários e problemas na interoperabilidade entre sistemas.

Um dos quesitos para se estimular a disponibilização dos dados de pesquisa pelos pesquisadores é a identificação de seus produtores (os pesquisadores) e isso se torna mais eficiente quando se utilizam dados conectados como identificadores persistentes, como o uso do ORCID e Lattes por meio de URI. Esta recomendação também é válida ao se utilizar vocabulários controlados para assuntos e tipologia documental, sendo de extrema importância as iniciativas do COAR e AGRAVOC de disponibilizar vocabulário controlado em formato de tesouro ou ontologia.

Assim, recomenda-se que novas investigações sejam realizadas com o intuito de se desenvolver modelos para a construção de um conjunto documental com o uso padronizado de esquemas de metadados, a aplicação de dados abertos e conectados, o compartilhamento de experiências existentes entre os diversos repositórios, além do trabalho cooperativo entre entidades de pesquisa.

Não se pode deixar de mencionar as dificuldades e as limitações que esta pesquisa teve, pois, ao se compartilhar as dificuldades enfrentadas, poderá auxiliar outros pesquisadores que passem pelas mesmas situações.

Primeiramente, ao se trabalhar com uma certa quantidade de planilhas, ao se coletar os dados, exige-se maior rigor e domínio do *software*, caso contrário, se incorrerá em erros que podem prejudicar o resultado da análise, portanto, é recomendável o aperfeiçoamento para a utilização dessas ferramentas. Uma segunda questão, ao se coletar e analisar informações em páginas na *internet*, deve-se ter em mente que são ambientes dinâmicos e em constante modificação, o que causou, nesta pesquisa, certo retrabalho, pois era necessário voltar aos ambientes para confirmar alguma informação e acabou-se por encontrar um ambiente modificado.

Por último, em consonância com as duas questões apresentadas, ao se coletar uma quantidade considerável de dados em um ambiente dinâmico, é necessário um trabalho mais objetivo e ágil, para se fazer em um menor período possível, evitando que ocorram distorções entre coletas ou encontrar páginas que não estejam disponíveis em determinado momento.

REFERÊNCIAS

ALVES, R. C. V.; SANTOS, P. L. V. A. C. **Metadados no domínio bibliográfico**. Rio de Janeiro: Intertexto, 2013.

ANDRADE, M. C. **Interoperabilidade semântica entre repositórios digitais científicos**: um perfil de aplicação de metadados baseado nas necessidades de informação em Oncologia. 2018. 457 f. Tese (Doutorado em Tecnologias e Sistemas de Informação) - Departamento de Sistemas de Informação, Escola de Engenharia, Universidade do Minho, Guimarães. 2018.

ARRAIZA, P. M.; LIBERATORE, G.; VIDOTTI, S. A. B. G. Publicação científica ampliada: desafios desde as ciências sociais e humanidades. *In*: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 19., 2018, Londrina. **Anais** [...] Londrina: UEL, 2018. Disponível em: http://enancib.marilia.unesp.br/index.php/XIX_ENANCIB/xixenancib/paper/viewFile/1260/1705. Acesso em: 25 jul. 2021.

AVENTURIER, P.; ALENCAR, M. C. F. Os desafios dos dados de pesquisas abertos. **RECIIS – Revista Eletrônica de Comunicação, Informação e Inovação em Saúde**, v. 10, n. 3, p. 1-19, jul./set. 2016. Disponível em: <https://www.reciis.icict.fiocruz.br/index.php/reciis/article/view/1069>. Acesso em: 24 jan. 2021.

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

BORGMAN, C. L. **Big data, little data, no data**: scholarship in the networked world. Cambridge: MIT Press, 2015.

BRÄSCHER, M. CAFÉ, L. Organização da Informação ou Organização do Conhecimento? *In*: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 9., 2008, São Paulo. **Anais** [...] São Paulo: USP, 2008. Disponível em: <http://repositorios.questoesemrede.uff.br/repositorios/handle/123456789/809>. Acesso em: 14 jan. 2021.

BROOKS, T. A. Where is meaning when from is gone? Knowledge representation on the web. **Information Research**, v. 6, n. 2, jan. 2001. Disponível em: <http://www.informationr.net/ir/6-2/paper93.html>. Acesso em:

BURKE, P. Problemas causados por Gutenberg: a explosão da informação nos primórdios da Europa moderna. **Estudos Avançados**, São Paulo, v. 16, n. 44, p. 173-185, abr. 2002. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0103-40142002000100010&lng=en&nrm=iso. Acesso em: 13 jan. 2021.

CAFÉ, L. M. A.; SALES, R. Organização da informação: conceitos básicos e breve fundamentação teórica. *In*: ROBREDO, J.; BRÄSCHER, M. (org.). **Passeios pelo bosque da informação**: estudos sobre representação e organização da informação e do conhecimento. Brasília: IBICT, 2010. p. 115-129.

CALLAGHAN, S. Research Data Publication: moving beyond the metaphor. **Data Science Journal**, v. 18, n. 1, p. 2-7, 2019. Disponível em: <https://datascience.codata.org/articles/10.5334/dsj-2019-039>. Acesso em: 19 jul. 2021.

CAPLAN, P. **Metadata fundamentals for all librarians**. Chicago: American Library Association, 2003.

CATARINO, M. E. **Integração das folksonomias aos metadados**: identificação de novos elementos de metadados como contributo para a descrição de recursos em repositórios. 2009. 232 f. Tese (Doutorado em Tecnologias e Sistemas de Informação) - Departamento de Sistemas de Informação, Escola de Engenharia, Universidade do Minho, Guimarães, 2009.

CÔRTEZ, P. L. Considerações sobre a evolução da ciência e da comunicação científica. *In: Comunicação e produção científica*: contexto, indicadores e avaliação. POBLACION, D. A.; WITTER, G. P.; SILVA, J. F. M. (Org.) São Paulo: Angellara, 2006.

CROSAS, M.; KING, G.; HONAKER, J.; SWEENEY, L. Automating open science for big data. **The Annals OF American Academy of Political and Social Science**, n. 659, p. 260-273, maio 2015. Disponível em: <https://gking.harvard.edu/publications/automating-open-science-big-data>. Acesso em: 14 ago. 2021.

DATA CITE Schema. **DataCite Metadata Schema**. Hanover: DataCite. 201-. Disponível em: <https://schema.datacite.org/> Acesso em: 02 fev. 2022.

DIAS, E. W. Organização do conhecimento no contexto de bibliotecas tradicionais e digitais. *In: NAVES, M. M. L.; KURAMOTO, H. Organização da informação*: princípios e tendências. Brasília: Briquet de Lemos, 2006.

DIAS, E. W.; NAVES, M. M. L. **Análise de assunto**: teoria e prática. 2.ed. Brasília: Briquet de Lemos, 2013.

FREIRE, G. H. Ciência da informação: temática, histórias e fundamentos. **Perspect. Ciênc. Inf.**, Belo Horizonte, v. 11 n.1, p. 6-19, jan./abr. 2006. Disponível em: <https://www.scielo.br/pdf/pci/v11n1/v11n1a02.pdf>. Acesso em: 04 jan. 2021.

FUJITA, M. S. L.; TOLARE, J. B. Análise do controle de vocabulários em repositórios institucionais brasileiros. *In: BARROS, T. H. B.; TOGNOLI, N. B. (org.) Organização do Conhecimento responsável*: promovendo sociedades democráticas e inclusiva. Belém: UFPA, 2019.

FUNDAÇÃO OSWALDO CRUZ. **Glossário da ciência aberta**. [20--]. Disponível em: <https://portal.fiocruz.br/glossario-da-ciencia-aberta>. Acesso em: 14 jan. 2021.

GARVEY, W. D. **Communication**: the essence of science. Oxford: Pergamon Press, 1979.

GIL, A. C. **Métodos e técnicas de pesquisa social**. 6. ed. São Paulo: Atlas, 2009.

GRACIO, J. C. A. **Metadados para a descrição de recursos da Internet: o padrão Dublin Core, aplicações e a questão da interoperabilidade**. 2002. 127 f. Dissertação (Mestrado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2002.

HJØRLAND, B. Theories of knowledge organization - Theories of knowledge. **Knowledge Organization**. v. 40, n. 3, p. 169-181. 2013. DOI: 10.5771/0943-7444-2013-3-169. Acesso em: 14 jan. 2021.

IBEKWE-SANJUAN, F.; BOWKER, G. C. Implications of big data for knowledge organization. **Knowledge Organization**, v. 44, n. 3, p. 187-198. 2017. Disponível em: <https://hal.archives-ouvertes.fr/hal-01489030/>. Acesso em: 16 jan. 2021.

ISOTANI, S.; BITTENCOURT, I. I. **Dados abertos conectados**. São Paulo: Novatec, 2015. Online. Disponível em: <https://ceweb.br/livros/dados-abertos-conectados/>. Acesso em: 28 dez 2021.

IVANOVIC, D.; SCHMIDT, B.; GRIM, R.; DUNNING, A. **FAIRness of repositories & their data: a report from LIBER's Research Data Management Working Group**. Zenodo, 2019.

KON, F. Ciência aberta, dados abertos e código aberto. **Computação Brasil**, Porto Alegre, n. 22, p. 22-60, jul. 2013. Disponível em: <http://www.ime.usp.br/~kon/papers/ComputacaoBrasilKon2013.pdf>. Acesso em: 25 jan. 2021.

LIMA, G. A. Organização e representação do conhecimento e da informação na web: teorias e técnicas. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 25, n. esp., p. 57-97, fev. 2020. Disponível em: <http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/4299>. Acesso em: 01 jun. 2021.

LIMA, F. R. B.; SANTOS, P. L. V. A. C.; SANTARÉM SEGUNDO, J. E. Padrão de metadados no domínio museológico. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 21, n. 3, p. 50-69, jul./set., 2016. Disponível em: <http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/2639>. Acesso em: 14 ago. 2021.

LOURENÇO, C. A.; ZAFALON, Z. R.; LOPES, A. T. Catalogação e metadados: reflexões sobre concepções, perspectivas e tendências. *In*: SIMÕES, M. G; LIMA, G. A. (org.). **Do tratamento à organização da informação: reflexões sobre concepções, perspectivas e tendências**. Coimbra: Imprensa Universidade de Coimbra, 2020. p. 83-118.

MAIMONE, G. D.; SILVEIRA, N. C.; TÁLAMO, M. F. G. M. Reflexões acerca das relações entre representação temática e descritiva. **Informação & Sociedade: estudos**, João Pessoa, v. 21, n. 1, p. 27-35, jan./abr. 2011. Disponível em: <https://periodicos.ufpb.br/ojs2/index.php/ies/article/view/7367>. Acesso em: 01 jun. 2021.

MEADOWS, A. J. **A comunicação científica**. Brasília: Briquet de Lemos, 1999.

MIRANDA, R. C. R. O uso da informação na formulação de ações estratégicas pelas empresas. **Ciência da Informação**, Brasília, v. 28, n. 3, p. 286-292, dez. 1999. Disponível em: <http://www.scielo.br/pdf/ci/v28n3/v28n3a6.pdf>. Acesso em: 24 jan. 2021.

MONTEIRO, F. S. **Organização da informação em repositórios digitais institucionais com ênfase na descrição física e descrição temática**. 2008. 199f. Dissertação (Mestrado em Ciência da Informação) – Universidade de Brasília, Brasília, 2008.

MULLER, S. P. M. A comunicação científica e o movimento de acesso livre ao conhecimento. **Ciência da Informação**, Brasília, v. 35, n. 2, p. 27-38, maio/ago, 2006.

OLIVEIRA, A. C. S. Publicações científicas ampliadas. *In*: SHINTAKU, M.; SALES, L. (org.). **Ciência aberta para editores científicos**. São Paulo: ABEC, 2019. p. 103-107.

OPEN DEFINITION. **The Open Definition**. 2021. [online]. Disponível em: <http://opendefinition.org/>. Acesso: 27 dez. 2021

OTLET, P. **Tratado de documentação**: o livro sobre o livro: teoria e prática. Brasília: Briquet de Lemos, 2018.

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, p.1-10, nov. 2013. Disponível em: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3817176/>. Acesso em: 05 jan. 2021.

PRÍNCIPE, P.; SCHIRRWAGEN, J. OpenAIRE guidelines for data source managers: aiming for metadata harmonization. *In*: CERN WORKSHOP ON INNOVATIONS IN SCHOLARLY COMMUNICATION, 9., 2015, Genebra. **Atas [...]**. Genebra: CERN, 2015. Disponível em: <http://hdl.handle.net/1822/35655>. Acesso em: 02 fev. 2022.

RIBEIRO, C. J. S. Modelo de Maturidade para Repositórios Digitais: um caminho para sua adoção na gestão de dados de pesquisa. **Liinc em Revista**, v. 15, n. 2, p. 224-243, nov. 2019. Disponível em: <http://revista.ibict.br/liinc/article/view/4816>. Acesso em: 16 jan. 2021.

RÍOS HILARIO, A. B.; GIL URDICIÁIN, B. Los datos bibliotecarios en la nube de datos: análisis de los datasets GLAM presentes en el LOD cloud diagram. **Scire**, Zaragoza, v. 21, n. 2, p. 35-47, jul./dez. 2015. Disponível em: <https://ibersid.eu/ojs/index.php/scire/article/view/4231/3811>. Acesso em: 4 jan. 2022.

RÍOS-HILARIO, A.; MARTÍN-CAMPO, D.; FERRERAS-FERNÁNDEZ, T. Linked data y linked open data: su implantación en una biblioteca digital. El caso de Europeana. **Profesional de la Información**, [S. l.], v. 21, n. 3, p. 292–297, maio/jun. 2012. Disponível em: <https://revista.profesionaldelainformacion.com/index.php/EPI/article/view/epi.2012.may.10>. Acesso em: 5 ago. 2021

RIVA, P.; LE BŒUF, P.; ŽUMER, M. **A conceptual model for bibliographic information**. Haia: IFLA, 2017.

RODRIGUES, F. A.; SANT'ANA, R. G. C. Publicação ampliada no contexto de teses e dissertações. **Informação & Tecnologia**, Marília/João Pessoa, v.3, n.1, p. 4-26, jan./jun. 2016.

RODRIGUES, A. A.; DIAS, G. A. Estudos sobre visualização de dados científicos no contexto da data Science e do Big Data. **Pesq. Bras. em Ci. da Inf. e Bib.**, João Pessoa, v. 12, n. 1, p. 219-228, 2017.

SAGAN, C. **Cosmos**. Rio de Janeiro: F. Alves, 1982.

SALES, L. F.; SOUZA, R. F.; SAYÃO, L. F. Publicação ampliada: um novo modelo de publicação científica voltada para os desafios de uma ciência orientada por dados. *In*: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 15., Belo Horizonte. **Anais [...]** Belo Horizonte: UFMG, 2014. Disponível em:

http://repositorios.questoesemrede.uff.br/repositorios/bitstream/handle/123456789/3182/2014_GT7-CO_18.pdf?sequence=1. Acesso em: 25 jul. 2021.

SALGADO, D. M. **O controle de autoridade sob a norma RDA**: análise da aplicação e implicações na construção de registros de autoridade. 2015. Dissertação (Mestrado em Cultura e Informação) - Escola de Comunicações e Artes, Universidade de São Paulo, São Paulo, 2015. Disponível em: <https://www.teses.usp.br/teses/disponiveis/27/27151/tde-19012016-105319/pt-br.php>. Acesso em: 10 mar. 2022.

SANTAREM SEGUNDO, J. E. **Representação iterativa**: um modelo para repositórios digitais. 2010. 224 f. Tese (Doutorado) - Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2010. Disponível em: https://www.marilia.unesp.br/Home/Pos-Graduacao/CienciadaInformacao/Dissertacoes/santaremsegundo_je_do_mar.pdf. Acesso em: 06 jan. 2022.

SANTOS, R. F.; NEVES, D. A. B.; SOUZA, E. D. A organização do conhecimento como domínio de estudo da Ciência da Informação: uma reflexão a partir dos aspectos epistemológicos. *In*: BARROS, T. H. B.; TOGNOLI, N. B. (org.). **Organização do conhecimento responsável**: promovendo sociedades democráticas e inclusivas. Belém: UFPA, 2019. p. 95-104.

SANTOS, P. L. V. C., SANT'ANA, R. C. G. Camadas de representação de dados e suas especificidades no cenário científico. *In*: DIAS, G. A.; OLIVEIRA, B. M. J. F. (org.) **Dados científicos**: perspectivas e desafios. João Pessoa: UFPB, 2019. p. 53-66.

SAYÃO, L. F.; SALES, L. F. Curadoria digital e dados de pesquisa. **AtoZ: novas práticas em informação e conhecimento**, Curitiba, v. 5, n. 2, p. 67-71, dec. 2016a. Disponível em: <https://revistas.ufpr.br/atoz/article/view/49708>. Acesso em: 04 jan. 2021.

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

<http://www.uel.br/revistas/uel/index.php/informacao/article/view/27939>. Acesso em: 04 jan. 2021.

SAYÃO, L. F.; SALES, L. F. A ciência invisível: por que os pesquisadores não publicam seus resultados negativos? **Informação & Informação**, Londrina, v. 25, n. 4, p. 98-116, out./dez. 2020. Disponível em:

<http://www.uel.br/revistas/uel/index.php/informacao/article/view/40016>. Acesso em: 04 jan. 2021.

SERRA, L. G.; ANDRADE, L. C. R. Sobre autoridades e identificadores únicos. *In*: SEMINÁRIO NACIONAL DE BIBLIOTECAS UNIVERSITÁRIAS, 20., 2018, Salvador. **Anais** [...]. Salvador: UFBA, 2018. Disponível em:

<http://repositorio.febab.org.br/items/show/5625>. Acesso em: 10 mar. 2022.

SILVA, F.C.C.; SILVEIRA, L. O ecossistema da Ciência Aberta. **Transinformação**, Campinas, v. 31, e190001, 2019. <http://dx.doi.org/10.1590/2318-0889201931e190001>. 16 jan. 2021.

SILVEIRA, N. C. Um diálogo sobre instrumentos de representação descritiva. *In*: ZAFALON, Z. R.; DAL'EVEDOVE, P. R. (org.). **Perspectivas da representação documental: discussão e experiências**. São Carlos: CPOI, 2017. p. 177-191.

SIMIONATO, A. C. Família FR, Bibframe e linked data: integração de dados bibliográficos. *In*: ZAFALON, Z. R.; DAL'EVEDOVE, P. R. (org.). **Perspectivas da representação documental: discussão e experiências**. São Carlos: CPOI, 2017. p. 211-226.

SOUSA, J. L.; RAMALHO, R. A. S. SKOS para vocabulários controlados. **Tendências da Pesquisa Brasileira e Ciência da Informação**, São Paulo, v. 13, n. 1, 2020. Disponível em: <https://revistas.ancib.org/index.php/tpbci/article/view/513>. Acesso em: 1 jun. 2021.

SOUZA, F. C. **Organização do conhecimento na sociedade**. Florianópolis: UFSC, 1998.

TORINO, E.; ROA-MARTÍNEZ, S. M.; VIDOTTI, S. A. B. G. Dados de pesquisa: disponibilização ou publicação? *In*: SHINTAKU, M.; SALES, L. F.; COSTA, M. (org.). **Tópicos sobre dados abertos para editores científicos**. Botucatu, SP: ABEC, 2020. p.183-201.

WEITZEL, S. R. E-prints: modelo da comunicação científica em transição. *In*: **Preparação de revistas científicas: teoria e prática**. FERREIRA, S. M. S. P.; TARGINO, M. G. (org.). São Paulo: Reichmann & Autores, 2005.

ZENG, M. L. Interoperability. **Knowledge Organization**, v. 46, n. 2, p. 122-146. 2019. Disponível em: https://www.ergon-verlag.de/isko_ko/downloads/ko_46_2019_2_d.pdf. Acesso em: 25 ago. 2021.